ABSTRACT

PRABHAKAR, VARSHA. A Method of Optimizing Pharmaceutical Drug Transport Routing to Minimize Diversion and Counterfeiting. (Under the direction of Dr. Julie Ivy, Dr. Anita Vila-Parrish).

This thesis describes a method to obtain the best route for transporting pharmaceuticals with the objective of reducing costs associated with diversion and counterfeiting by country. The method described first quantifies the relationship between diversion, counterfeiting and 16 independent variables by using linear regression methods (Ordinary least squares, Ridge regression, Lasso regression), nonlinear regression method (Random forests) and classification (Ordinal logistic regression). The relationship obtained is then used to predict the quantity of diversion and occurrence of counterfeiting. The expected costs of diversion and counterfeiting calculated for these predictions along with costs associated with transportation are used to create an expected cost matrix for all possible routes. Finally, Dijkstra's algorithm is used to find the best route for transportation. The regression and classification analysis revealed a linear relationship between diversion and counterfeiting. Out of the 16 independent variables used, it was found that diversion shared a relationship with the interaction effects of counterfeit incident index and population, counterfeit incident index and cargo theft index. Counterfeiting was found to have a linear relationship with the corruption perception index and diversion in a country. The model was tested using data from the distribution network of a logistics company - Agility, in Southeast Asia. The results and sensitivity analysis revealed that the performance of the model relied on the quality of data used for regression analysis. The limitations due to the unavailability of data and suggestions for future work to improve this model are also discussed.

A Method of Optimizing Pharmaceutical Drug Transport Routing to Minimize Diversion and Counterfeiting

by
Varsha Prabhakar

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Industrial Engineering

Raleigh, North Carolina

2016

APPROVED BY:

Dr. Julie Ivy Co - Chair	Dr. James Wilson Minor

Dr. Anita Vila-Parrish
Co - Chair

BIOGRAPHY

Varsha Prabhakar was born on August 8, 1989 in Mysore, India. She obtained a Bachelor of Engineering degree in Industrial and Production Engineering from Visveswaraya Technological University at People's Education Society College of Engineering, Mandya. She then pursued a Master of Science degree in Industrial Engineering at North Carolina State University under the guidance of Dr. Julie Ivy.

ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Julie Ivy, for her support, and patience while mentoring me. Her unwavering passion and determination to solve difficult problems inspired me to continue working towards the completion of my thesis by learning from my mistakes and missteps.

I would also like to thank my committee, Dr. James Wilson and Dr. Anita Vila-Parrish for evaluating my thesis and providing valuable critiques.

I want to thank my parents, family, friends and Abhishek for believing in me and supporting me through all the ups and downs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3PL	Third Party Logistics
ADC	Automated Dispensing Cabinets
CDC	Centers for Disease Control and Prevention
CII	Counterfeit Incident Index
CMS	Centre for Medicare and Medicaid Services
CPI	Corruption Perception Index
CTI	Cargo Theft Index
DEA	Drug Enforcement Agency
EMEA	Europe, The Middle East & Africa
FDA	U.S. Food and Drug Administration
GDP	Gross Domestic Product
INTERPOL	International Criminal Police Organization
IPRI	Intellectual Property Rights Index
LPI	Logistics Performance Index
MAR	Missing at Random
NIDA	National Institute On Drug Abuse
OLS	Ordinary Least Squares
RFID	Radio Frequency Identification
SCIC	Supply Chain Intelligence Center
SIS	Sure Independence Screening
TAPA	Transport Asset Protection Association
UN COMTRADE	United Nations Commodity Trade Statistics Database
UNODC	United Nations Office on Drugs and Crime
WHO	World Health Organization
WHOSIS	WHO Statistical Information System
WIPO	World Intellectual Property Organization

LIST OF NOTATIONS

N	Total number of survey responses
n_i	Sample size of country i in N
$\bar{Z}_j$	Sample mean for core component j
σ_j	Sample standard deviation for core component j
z_{ij}	Survey result of core component j for country i
$\widehat{\boldsymbol{\beta}}^{OLS}$	OLS estimate of coefficients
$\boldsymbol{X}$	Design matrix
y	Vector of diverted drug seizure estimates
$\widehat{\boldsymbol{\beta}}^{R}$	Ridge estimate of coefficients
$\boldsymbol{\lambda}$	Ridge regression tuning parameter such that
$\boldsymbol{X^{test}}$	Test Data set
α_j	Intercepts
β	Coefficients
C	Cost of one-unit of drug in the given region in USD
S	Estimated counterfeit incident index
n	Number of incidents per estimated counterfeit incident index~ U (100,999999)
q	Average amount of counterfeit copies detected per incident reported
d(m)	Distance of transport as a function of mode of transport
T(m)	Cost of transportation as a function of mode of transport
G = (V, A)	Weighted graph representing the routing problem
V (G)	Set of vertices representing countries
A(G)	Set of all arcs connecting adjacent vertices
$w(v_{k-1,k})$	Cost of entering country v_k from country v_{k-1}
O	Set of countries to which minimum cost route from origin u=v_0 is known
Q	Set of countries in queue for which minimum cost route from origin u is unknown
t (v_k)	Length of least expensive route found yet from origin u to country v_k

1 Introduction

The previous decades have seen an increased interest in securing the pharmaceutical supply chain to prevent drug crimes. Many steps taken to reduce research, manufacturing and distribution costs have resulted in the expansion of the supply chain, thus exposing it and giving more room for crime. Globalization of the pharmaceutical supply chain has reduced visibility and has made it easier for crimes such as drug diversion and counterfeiting to occur (1). The total number of pharmaceutical crimes reported by the Pharmaceutical Security Institute (PSI) has increased by 51.16% from 1986 incidents in 2011 to 3002 incidents in 2015 (2).

Major pharmaceutical companies have come together to create non- profit organizations such as the PSI and Rx 360 to collect and share data for enhancing supply chain security. According to PSI, they are "a not-for-profit, membership organization dedicated to protecting the public health, sharing information on counterfeiting of pharmaceuticals and initiating enforcement actions through appropriate authorities" (3). Rx 360 is a not-for-profit consortium led by volunteers from manufacturers and suppliers from the pharmaceutical and biotech industries. The purpose of Rx 360 is to enhance healthcare supply chain security with the mission of protecting patient safety by sharing information and developing processes related to the integrity of the supply chain (4).

The aim of this thesis is to develop a model that can quantify the risk of diversion and counterfeiting to arrive at the best possible route for reducing diversion and counterfeiting due to incidents of thefts, hijack, robbery and burglary along the distribution channels. Knowledge regarding factors which can indicate or contribute to diversion and counterfeiting can be used by the key players of the pharmaceutical supply chain towards improving the security and quality of the pharmaceutical products.

1.1 Pharmaceutical Supply Chain

The downstream pharmaceutical supply chain, i.e., the flow of products from manufacturers to patients is different from other supply chains. The product is seldom sold directly to the end-user in the pharmaceutical product distribution (5). Instead, these products are sold to wholesalers in bulk who in turn sell them at a lower profit margin to retailers (5). Consumers can access these products from the retailers.

1.1.1 Pharmaceutical Supply Chain in the United States

The key players of the U.S. pharmaceutical supply chain in the distribution process are:

- Manufacturers
- Distributors
- Hospitals or Pharmacies

Figures 1-1 and 1-2 display the key players in the pharmaceutical supply chain in the United States (U.S.) and the flow of products between them.

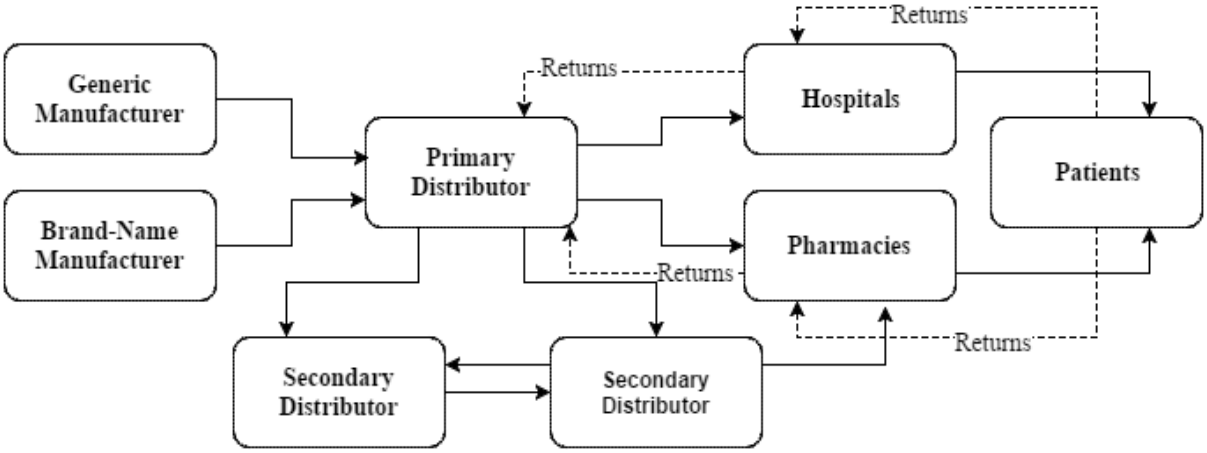

Figure 1-1 Commercial Pharmaceutical Supply Chain in United States

The role of each player is described below:

Manufacturers in the US manufacture brand-name drugs, generic drugs or both types of drugs. Brand-name drugs are patented drugs that are discovered and developed by a pharmaceutical company. These drugs are usually more expensive because of the time and money spent in researching, developing and testing them (6). Generic drugs sold at a discounted rate, are chemically identical to brand-name counterparts manufactured under the U.S. Food and Drug Administration's (FDA's) good manufacturing practice regulations for innovator products (7). Both types of manufacturers ensure patient safety by providing informational labeling for prescribers and consumers. They also use technologies such as bar-coding to track individual production lots (8).

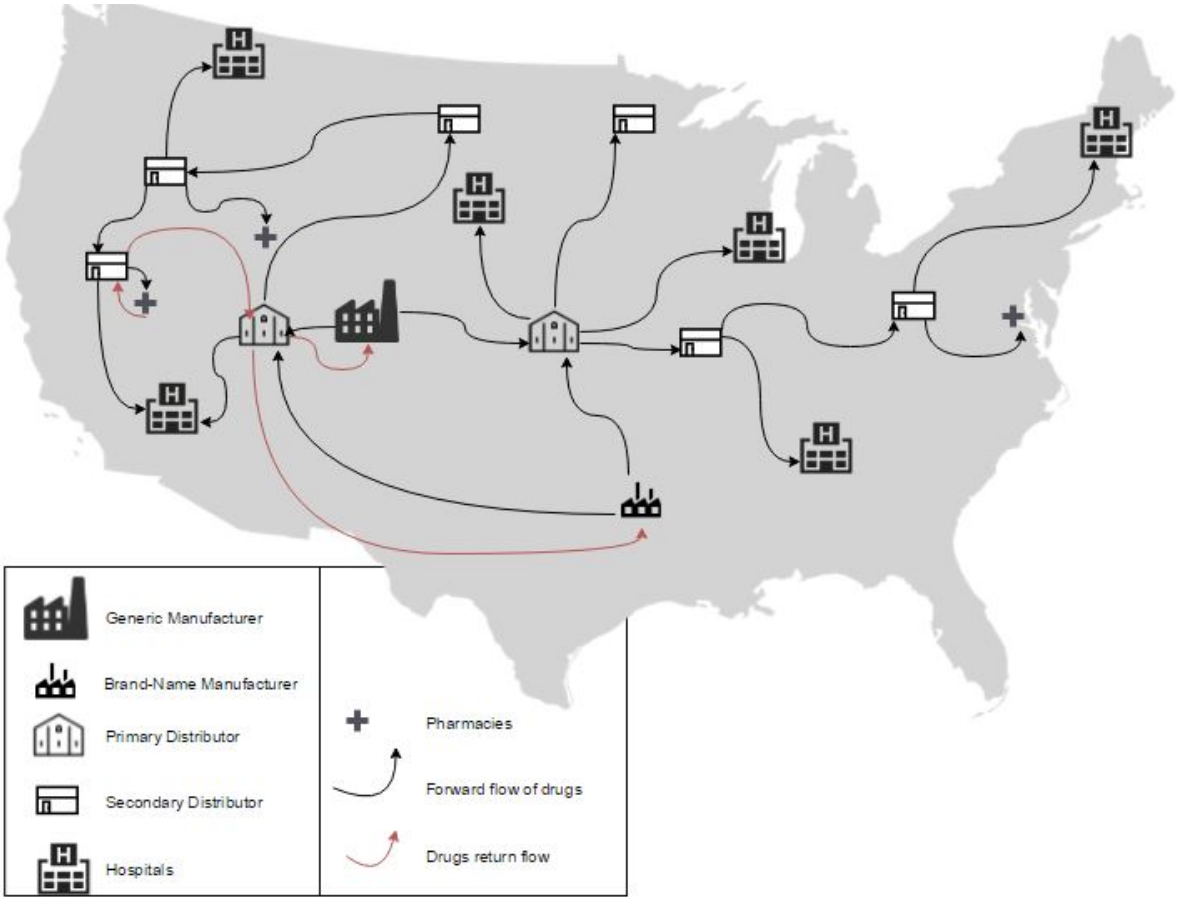

Figure 1-2 Pharmaceutical distribution in the United States

Primary distributors buy drugs from manufacturers in bulk, store them in their warehouses and distribute them to hospitals, pharmacies and other distributors (5). Primary distributors, also

referred to as wholesalers or Authorized Distributors of Record (ADR) enter into contracts with manufacturers and buy drugs from them directly (9). The U.S. Food and Drug Administration (FDA) defines ADR as "a distributor with whom a manufacturer has established an ongoing relationship to distribute the manufacturer's products". They can either be national or regional and are responsible for supplying up to 91% of prescription drugs sold in the United States (10). The US has three major national primary distributors, AmerisourceBergen Corp (ABC), Cardinal Health, Inc. (CAH), and McKesson Corporation (MCK). Together, these distributors are referred to as the "Big Three" and generated a total revenue of $327.7 Billion in 2014 (11).

According to the FDA, secondary distributors specialize in buying and selling discounted drugs. They are "distinguished by their willingness to risk substantial capital in buying and trading discounted drugs" (12). These distributors do not have a contract or an agreement with the manufacturers to purchase and resell their drugs. They usually purchase drugs from primary distributors, other secondary distributors and sometimes directly from the manufacturing unit (13).

Secondary distributors are considered to be the weakest links in the pharmaceutical distribution chain (9). Mainly because of their repeated buying and selling of medicines for discounted rates to fulfil the market demand. However, they are still considered to be important in the pharmaceutical distribution chain. Secondary distributors can reach rural areas whose medical demands are not fulfilled by primary distributors (13). They may also supply medicines to primary distributors and retailers by procuring them from other distributors during any drug shortage (9).

1.1.2 Global Pharmaceutical Supply Chain with Contract Manufacturing Organizations

Contract manufacturing organizations (CMO) are companies which offer manufacturing services to pharmaceutical companies from smaller to larger volumes for clinical purposes or commercialization (14). According to Zhang *et al.* (2013), using CMOs located in emerging economies also provides pharmaceutical companies with options to "reduce cost, improve

speed, quality and flexibility, and adjust their organizational boundaries in response to external economic pressures" (15).

Using these CMOs located in emerging economies increases the length of the supply chain due to globalization and also adds more stake-holders to the network. The key stake-holders in the global pharmaceutical supply chain are:

- Outsourcer (Pharmaceutical Company)
- Contract Manufacturing Organization
- Distributor or Wholesalers
- Hospitals or Pharmacies

Boulaksil and Fransoo, (2010), describe the flow of products in the presence of a CMO in the global supply chain using two case studies. In these case studies, an outsourcer from Ireland has outsourced manufacturing to 2 CMOs, Contract Manufacturer 1 (CM1) is in the USA, Contract Manufacturer 2 (CM2) is in Germany. Both the CMOs use the same active ingredient (AI) but manufacture different types of products. A CM1 produces blistered tablets that are sent to the national warehouses belonging to the outsourcer in the USA (US), Netherlands (NL), Mexico (MX) and Brazil (BR). These are then packaged and sent to warehouses around the world (RW) either to be sold to customers directly or to pharmaceutical wholesalers. A CM2 produces vials of dissolved active ingredient which are sent to the outsourcer's production site in Ireland. Vials are sent from this production site to a national warehouse in the US which packs vials for Brazil and USA while the site in Ireland (IR) packs vials for the rest of the world (RW). These are then sold to wholesalers, hospitals or pharmacies (16). Figure 1-3 summarizes the flow of these products.

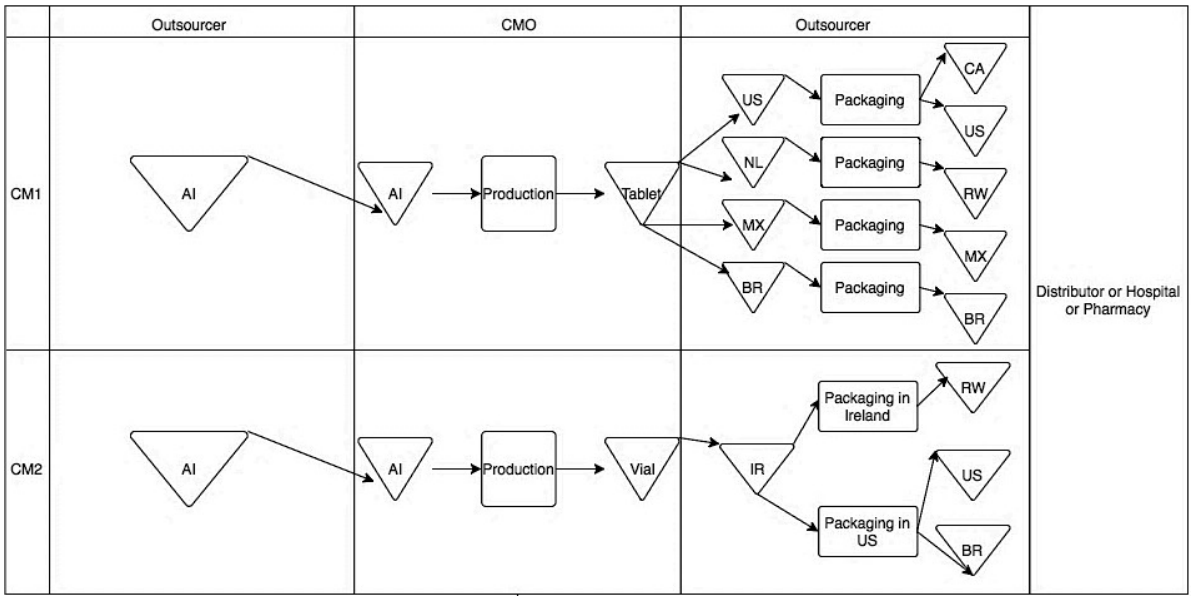

Figure 1-3 Flow of products in a global pharmaceutical supply chain with CMOs (Boulaksil and Fransoo, (2010))

Transportation of products between these stake-holders in the supply chain can be done in many ways. However, due to globalization many drug manufacturers are now relying on third party logistic providers (3PL providers) to transport pharmaceutical products (17). The 3PL providers specialize in transportation and warehousing services and usually provide additional services suitable for the manufacturer's or distributor's requirements. According to Global Market Insights, a market research and consulting company, the market worth of 3PL providers will increase from USD 733.69 Billion in 2014 to USD 1029.47 Billion by 2020 (18).

1.2 Diversion

According to the Center for Medicare and Medicaid Services (CMS), "Drug Diversion is the diversion of licit drugs for illicit purposes" (19). Diversion occurs when drugs are moved out of the supply chain towards illegal or non-medical uses. Since the drugs are genuine but are being sold/ procured illicitly, diversion is also referred to as gray market activity. This can occur at multiple points in the supply chain for reasons ranging from drug abuse to acquiring profit from differential pricing.

One of the major reasons for and indicators of diversion is pharmaceutical drug abuse. Drug Enforcement Agency (DEA) records indicate a 115% increase from 2004 to 2010 in the number of ER visits due to pharmaceutical drug abuse in North America (20). Abusers mainly target opioids/opiates such as oxycodone and hydrocodone products followed by psychotherapeutic drugs such as Citalopram, Quetiapine, Zolpidem, Alprazolam and Clonazepam (21). They obtain drugs either by diverting them themselves or by procuring them through groups or individuals who divert drugs and sell them without the need for prescription.

People who are responsible for diverting pharmaceutical drugs often receive minor punishments and fines when they are caught. They sell diverted drugs at a higher price to abusers. Thus, it is safe and lucrative for these people to divert pharmaceutical drugs instead of illicit drugs, i.e., "drugs under international control which are produced, trafficked and/or consumed illicitly" (22), which can result in a harsher punishment (23). Some of these diverters are drug abusers themselves who trade in pharmaceutical drugs for other illegal drugs (24) while, others stand to gain from differential pricing.

International pricing arbitrage is considered to be one of the vulnerabilities to diversion affecting global supply chain. Differential pricing can occur due to marketing policies adopted by countries and due to the lack of availability or access to pharmaceutical drugs (25). Since these diversions are motivated by economic factors, they are not just limited to diversion of opioids and psychotherapeutic drugs but extend to other lifesaving drugs as well. The Avastin Scandal in 2012 brought this issue to the forefront, where illegally diverted Avastin was sold at a price 20% lower than the actual price through a Canadian internet pharmacy (25).

Avastin, a drug used along with chemotherapy to treat different forms of cancer, is manufactured by Roche Holding AG, Switzerland. Fake batches of this drug consisting of salt, starch and other chemicals, but no active ingredients entered the legitimate supply chain in Turkey in 2012 (25). A Swiss distributor, unaware of the quality purchased this fake drug from Turkey through a Syrian middleman from Egypt and sold it to another distributor in Copenhagen (9). The drug was then sold under the company, Canada Drugs, to other companies (9).

Existence of secondary distributors allows for the exploitation of differential pricing to make profits through diversion. In many cases, these illegally diverted drugs get back into the legal supply chain to an authorized distributor of record through secondary distributors which happened in the Avastin case. Drugs can also be transferred directly to recipients by internet pharmacies such as RxNorth (26) who may falsify information regarding FDA approval for selling prescription drugs.

This uncertainty in the reliability of drugs received from secondary distributors has driven many group purchasing organizations and state boards of pharmacy to recommend wholesalers to procure drugs directly from manufacturers (13). However, secondary distributors are an essential part of the supply chain and thus make it difficult to remove them completely. Not all secondary distributors engage in diversion either, but there are overlaps between the gray market and secondary distributors. These distributors take advantage of situations such as drug shortage. For example, in 2011, shortage of Cytarabine, a chemotherapy agent used to treat Acute Myeloid Leukemia (AML) and non-Hodgkin Lymphoma (NHL) resulted in secondary distributors from southern Florida increasing the price by 7500% (27).

Many secondary distributors procure drugs in different ways and are not concerned about where the drugs come from. They build a network which adds to the complexity of the supply chain by inter -connecting licit and illicit sources which are not just limited to diverted drugs. There have been many instances of these secondary distributors gathering drugs closer to expiry from manufacturers at discounted rates (13). This overlap between gray market and secondary distributors also allows counterfeit drugs to penetrate the legal supply chain (25) (28).

The opportunity for lucrative deals to be made is the reason all stages of the supply chain including reverse logistics - the process of returning expired, banned, unused or recalled drugs to manufacturers or for waste disposal, are vulnerable to diversion (29). Figure 1-3 shows the movement of legitimate drugs and diverted drugs and illustrates points at which diversion can occur.

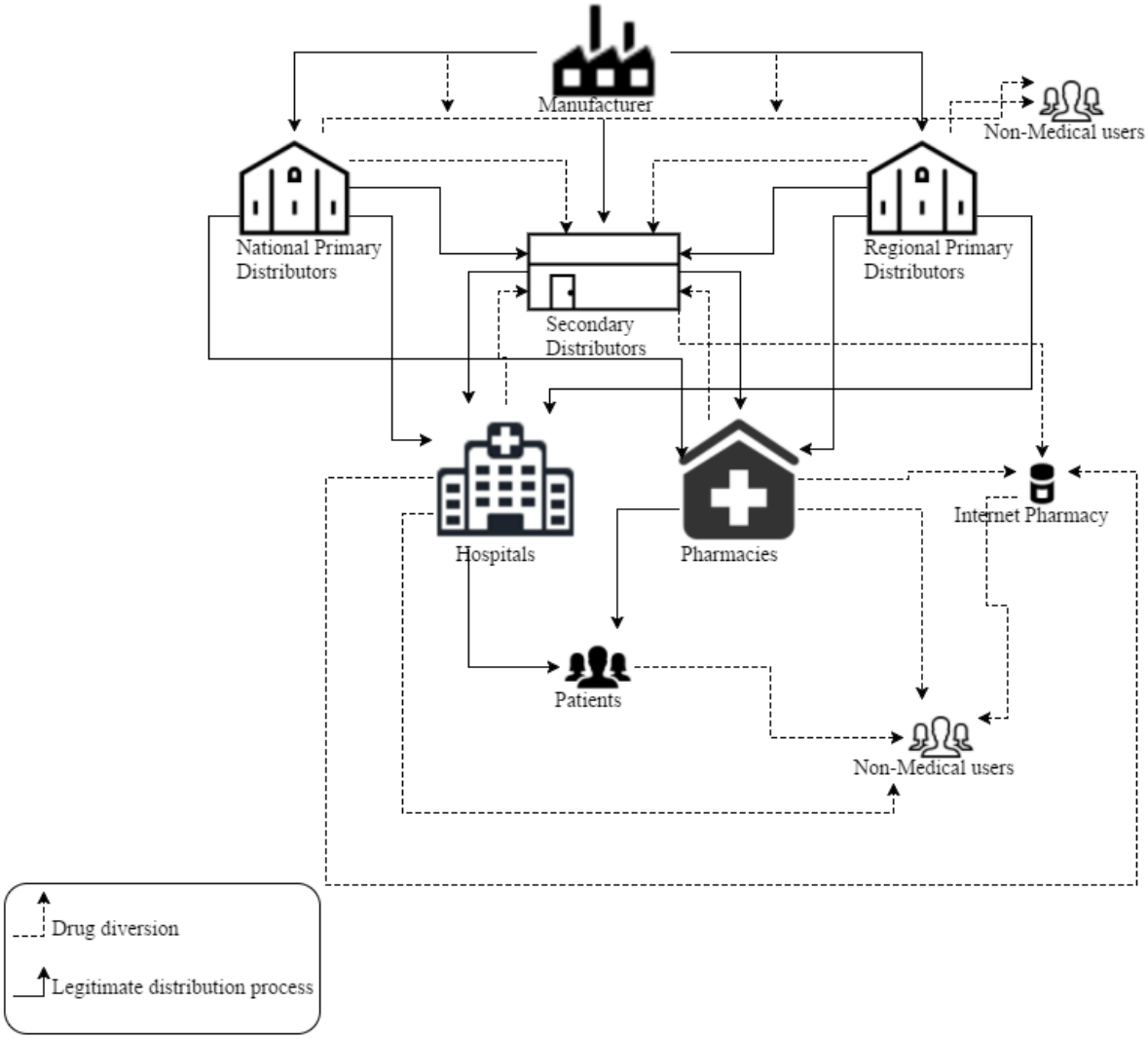

Figure 1-4 Diverted and legitimate drug flow channels

Table 1.1 describes the different ways in which drug diversion can occur at different points with examples. The scale of diversion may vary based on where the diversion occurs. This scale is determined based on where the diversion occurs and the possible ways in which these diverted drugs may be re-sold or used. Diversion on a global scale indicates that the diversion affected more than one country, diversion on a national scale indicates that the diversion affected the entire country. Similarly, regional scale indicates diversion affecting a region and local scale is restricted to a particular pharmacy, hospital or an individual. This shows how vulnerable the supply chain is at every point.

Table 1-1 Methods of Drug Diversion

Points in Supply Chain where Diversion can take place	**Method**	**Example**	**Scale of diversion**
Distribution	Cargo Theft	In May 2009, a gang hijacked pharmaceutical shipment worth USD 1.5 million in Mississauga, Canada (30).	National
	Secondary Distributors	Cumberland Distributors illegally obtained unused prescription drugs worth USD 58 million and sold them to pharmacies and hospitals in 2013 (31).	National
	Organized Crime	Italian mafia stole Herceptin and sold them along with counterfeits throughout Europe in 2014 (32).	Global
	Last mile thefts	In May 2015, a group of armed robbers posed as DEA, stopped a delivery van in Alabama and made away with drugs worth $108,000 (33).	Regional
Storage	Theft from Distribution Warehouse	Drugs worth USD 3 million were stolen from Exel Distribution Center's warehouse in 2009 (34).	National
	Theft from Hospital Warehouse	In 2006, a registered nurse stole narcotics using stolen passwords from Automated Dispensing Cabinets (ADCs) while they were undergoing system upgrades (35).	Local

Table 1-1 continued

Pharmacies/Hospitals/Doctors office	At the counter	A pharmacist forged prescriptions to obtain drugs **(36)**.	Local
	Patient Care area	A radiology technician diverted unused fentanyl syringes and replaced them with saline solution (37).	Local
	At disposal	A healthcare worker at a Mayo clinic was discovered retrieving narcotics from waste containers in 2010 (37).	Local
Reverse Logistics		A police chief was obtaining prescription drugs meant for disposal by posing as an intermediary between police and DEA to abuse those drugs (38).	Regional

1.3 Counterfeiting

The Centers for Disease Control and Prevention (CDC) defines counterfeit drugs as medicines manufactured with incorrect or harmful ingredients, packaged and labeled to look genuine. Unlike diversion, most of the target group for counterfeit drugs are people who require drugs for medical use. Counterfeits might have too much, very little or none of the active ingredient, which make them dangerous and ineffective. INTERPOL states that around 200,000 children die every year in Africa because of counterfeit anti-malaria medication (39). It is estimated that 10% of drugs being sold worldwide are counterfeits (40). This situation is more prevalent in developing nations such as Asian countries where the number of counterfeits sold range from 30% to 60%. Industrialized countries have more secure pharmaceutical supply chains and have a low estimate of 1% occurrence of counterfeits (40).

The main motivator for counterfeiting is that of economic gain. The bulk of these counterfeits are manufactured in countries like China and India (40). These countries are major exporters of pharmaceutical drugs to African countries (41). Lifesaving drugs against malaria, tuberculosis and HIV/AIDS are the most frequently counterfeited medicines in Africa (42) (43).

Counterfeits gain easy access to the legitimate supply chain when secondary distributors, pharmacists and hospitals are willing to risk the quality of drugs available at lower costs. Approximately 91% of drugs in USA are imported through primary distributors, the remaining 9% are imported through secondary distributors with an increased likelihood of counterfeits entering the legitimate supply chain (44). Existence of regional and secondary distributors increases the number of links in the supply chain. These distributors are also involved in repacking and labeling which adds to the risk (45).

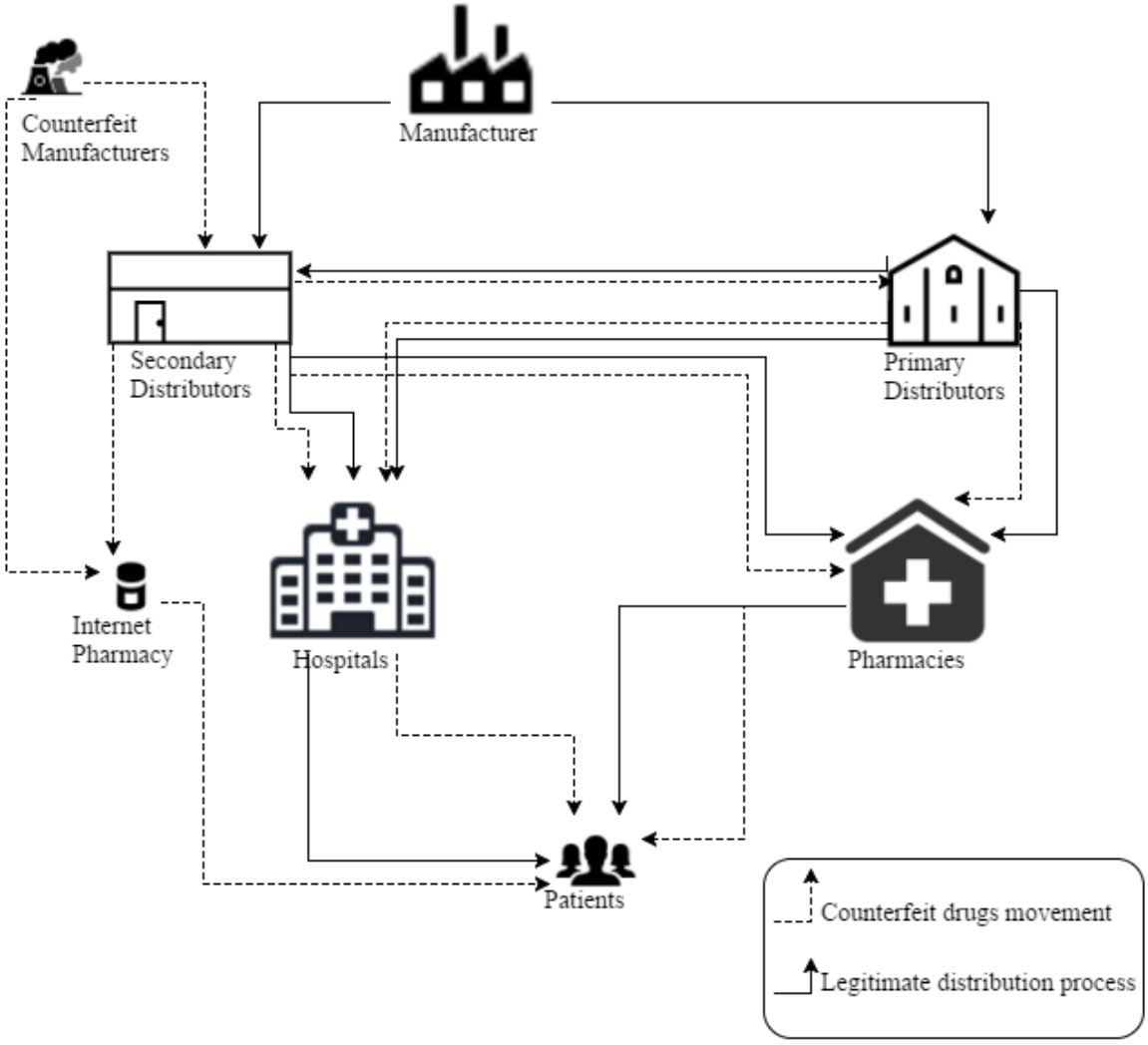

Figure 1-5 Counterfeit and legitimate drug flow channels

Counterfeit drugs can enter the legitimate supply chain at various points. Figure 1-5 shows the movement of legitimate and counterfeit drugs illustrating points at which counterfeits can enter the legitimate supply chain. It is difficult to trace the flow of counterfeits when they are detected. The Avastin scandal which gained so much attention remains very complicated even now since the point of origin is still unknown (46) . In many cases, counterfeits are not reported because they go undetected, especially in cases like dilution where the active ingredient is present, but in a slightly lower quantity (47).

Illegally diverted drugs that are brought back into the legitimate supply chain add to the risk of counterfeiting. The Avastin scandal shows how authorized channels in the supply chain were procuring diverted drugs at a lower cost, which made these authorized channels vulnerable to counterfeits (46). Diverted drugs can also be tampered with and re-introduced as counterfeits to the supply chain (47). These instances indicate a connection between diversion and counterfeiting during the distribution stage of the supply chain before the drugs reach health care personnel through methods such as cargo crimes.

Pharmaceutical cargo crime has been on the rise. The Transport Asset Protection Association (TAPA) of Europe recorded 15 incidents in 2014, 12 incidents in 2015 and 12 incidents by July of 2016 (48). Despite the dip in number of incidents in 2015, the average value of pharmaceutical theft per incident showed a steep increase (48). This has been a cause for concern due to the economic and patient safety impacts. A pharmaceutical cargo theft might result in the recall of an entire consignment, which can affect the manufacturing company economically. These stolen drugs may also be re-introduced into the supply chain by criminals who are unaware of, or not care about temperature and humidity requirements, thus making the drugs useless or harmful for patients (49).

Many pharmaceutical companies have formed non-profit organizations to collect data regarding these incidents and to assess the risks that make their supply chain vulnerable to diversion and counterfeits. Supply chain security is very important, especially due to the number of contract manufacturing organizations in developing countries. The existence of these organizations might lower manufacturing cost considerably but they increase number of supply chain links and are more vulnerable to counterfeiting. Most of these CMOs are in China and India which are the two major manufacturers of counterfeit drugs (50).

1.4 Literature Review

One of the solutions which has been discussed in the literature is that of increasing traceability of drugs through the supply chain (28). Literature suggests that technology can help make the

supply chain transparent through techniques like Holograms, Barcoding, Radio Frequency Identification (RFID). RFID uses radio waves to store serial numbers tagged with these drugs. These can be used to track and verify the authenticity and quality of drugs at all stages of the supply chain (51).

These tracking techniques have their own pros and cons. They are very useful since they enable supply chain members to track drugs at any point in the supply chain and allow the end user to trace the flow to the origin. But, most of these tracking codes are present on the packaging and thus lose their value in cases where the drugs are repackaged by secondary distributors (52). RFID is considered to be more efficient compared to other techniques (52), however RFID still remains an emerging idea in pharmaceutical logistics due to its cost of implementation and government policies which require a certain type of labelling and tracking practice for drugs (53)

Apart from tracking technology, suggestions have been made about routing cargo to increase safety while the drugs are in transit (54). However, this area has not been pursued in depth, mainly due to lack of availability of data regarding diversion and counterfeiting incidents. Organizations such as Rx-360 and the Pharmaceutical Security Institute have been recording any reported data regarding diversion or counterfeiting of drugs. FreightWatch International, a provider of logistic security services, uses active monitoring solutions to increase cargo visibility from origin to destination. They also provide a route analysis tool to track and maintain data on cargo thefts along different routes.

This thesis research explores routing based on regional risk analysis with the goal of addressing the issue of diversion and counterfeiting due to incidents of thefts, hijack, robbery and burglary along the distribution channels. The aim of this thesis is to understand factors which can affect the risk of and increase the likelihood of diversion and counterfeiting. This will especially be useful for manufacturers and distributors who can use this knowledge to reduce such incidents. This knowledge can also be used to assess the possibility of a drug procured through high risk routes being fraudulent or counterfeit.

Chapter 2 discusses the different regression and classification methods used to assess the possibility and amount of diversion and counterfeiting that can occur in countries around the world. This information is consolidated and used to determine optimal routes for transporting drugs at the global level. Results and findings from the thesis are presented in Chapter 3. Sensitivity analysis is used to test robustness of the model in Chapter 4. Chapter 5 contains the conclusion and opportunities for future work.

2 Methodology

The outline for this study includes data collection, predictive modeling and cost estimation, all of which serve as inputs to the routing model.

2.1 Data Collection and pre-processing

All of the data obtained were from open public sources. This process started by gaining an understanding of how and where diversion and counterfeiting can occur in the supply chain. This knowledge along with a study of factors motivating diversion and counterfeiting served as the foundation for data collection.

Literature has previously focused on diversion at the end of the supply chain, where drugs are diverted by healthcare providers, patients or through faked prescriptions (37) (24) (55) (56). However, this project focuses on diversion occurring during transportation of drugs. Thus, a greater emphasis was given on understanding the factors influencing both counterfeiting and diversion as well as factors which indicate a higher correlation with diversion and counterfeiting.

Table 2.1 presents the most commonly diverted prescription drugs according to the National Institute on Drug Abuse (NIDA). Diversion data related to these drugs were collected. Since there is no publicly available data regarding the actual amount of illegal diversion (57), data

from seizure reports of illegally diverted drugs mentioned in Table 2.1 was used as an estimate of the volume of drugs being diverted.

Table 2-1 Commonly Diverted Prescription Drugs

Drug Name	Trade Name	Type of medication
Alprazolam	Xanax	Benzodiazepine
Amphetamines	Adderall	Synthetic Psychoactive
Carisoprodol	Soma	Muscle relaxant
Diazepam	Valium	Benzodiazepine
Methylphenidate	Ritalin	Central Nervous System Stimulant
Oxycodone	OxyContin	Opioid
Fentanyl	Duragesic	Narcotic
Hydrocodone	Vicodin	Narcotic
Hydromorphone	Dilaudid	Opioid
Meperidine	Demerol	Opioid
Oxymorphone	Opana	Opioid

Pharmaceutical imports data was collected to compare the number of drugs diverted to the number of drugs imported into a country. This provided a baseline for understanding what percentage of the drugs being transported into/through a country was reported to be diverted. This however is not an exact indication of diversion. Many diversion incidents go unreported. So, the data collected here only serves as a lower bound. Sensitivity analysis was used to quantify the impact of this data limitation.

One of the major motivating factors for diversion is to provide prescription drugs to those who abuse them (20). Also, there have been discussions of a correlation between prescription drug

abuse and substance abuse (58) (59) (60). But, there are not many reliable estimates for the number of prescription drug abusers (57). Thus, data related to previous substance abuse (mainly opioids and alcohol) prevalence was used as a surrogate.

Different countries have different rules regarding punishing criminals involved in pharmaceutical drug offences. Illegal diversion of pharmaceutical drugs is generally considered harmless in relation to illicit drug trafficking and is thus less frowned upon. This results in insignificant punishments for the criminals involved (23). According to the World Drug Report (2012), this type of false security has resulted in a higher number of women abusing prescription drugs as compared to a higher number of men abusing illicit drugs. A survey conducted by the United Nations Office on Drugs and Crime (UNODC) in Afghanistan in 2009 (61), showed that women were twice as likely as men to abuse prescription drugs especially sedatives and tranquilizers. UNODC was established in 1997 to fight against illicit drugs and international crime. Data regarding the percentage of women and sex ratio was collected from the World Bank. According to the World Bank, sex ratio is the 5-year average of male births per female births (62). The number of women was determined as shown below:

$$\text{Number of Women} = \begin{cases} \%\, women \times population, if\ data\ collected = \%\, women \\ \frac{Total\ Population}{(1+Sex\ Ratio)}, if\ data\ collected = sex\ ratio \end{cases} \quad (2.0)$$

The Cargo Theft Index (CTI), was assumed to be representative of the risk of pharmaceutical cargo thefts within each country (30). This data was collected using a heat map by FreightWatch International's Global Cargo Threat Assessment Report (30) which indicated five levels of cargo theft risk. Generally, the risk of cargo theft may vary within a country, however for this study the data was collected by grouping the entire country under one risk level. Table 2.2 shows how the risk level was mapped to categories of the CTI.

Table 2-2 Risk level for Cargo theft index

Cargo theft index (CTI)	**Cargo theft risk level**
1	Low
2	Elevated
3	Moderate
4	High
5	Severe

The Logistics Performance Index (LPI) by the World Bank is a qualitative international trade tool which assesses countries based on customs performance, infrastructure quality, quality of logistics services and timeliness of shipments. All of this information is aggregated into a single index which increases the ease of comparison. A country with higher value of LPI indicates a more secure and safer region for transit of drugs. In the case of counterfeiting of drugs, LPI gives an estimate of the quality assurance standards of a country as well as the amount of security provided through customs to inhibit counterfeit drugs from entering the country.

According to the World Bank, raw data for LPI was collected by conducting a survey on six core logistics performance components shown in Table 2.3.

Table 2-3 Rating scale for the core logistics performance components

Core Component	**Rating scale**
Efficiency of customs and border clearance	Very low (1) to very high (5)
Quality of trade and transport infrastructure	Very low (1) to very high (5)
Ease of arranging competitively priced shipments	Very difficult (1) to very easy (5)
Competence and quality of logistics services	Very low (1) to very high (5)
Ability to track and trace consignments	Very low (1) to very high (5)
Frequency with which shipments reach consignees within scheduled or expected delivery times	Hardly ever (1) to nearly always (5)

Each survey respondent rated eight countries which were the most important trade markets of the country where the respondent was located. After 200 surveys, a data set was created by selecting countries using a Uniform Sampling Randomized (USR) technique, thus choosing countries at random but with non-uniform probabilities given by:

$$Probability\ of\ choosing\ a\ country\ i = \frac{N - n_i}{2N} \tag{2.1}$$

Where,

N = total number of survey responses

n_i = Sample size of country i in N

The data set created using USR was used as an input to principal component analysis (PCA) to obtain the LPI. Principal component analysis is a statistical dimensionality reduction technique. The input data set contained normalized survey results for the chosen countries. The survey results were normalized using the following formula:

$$\frac{\bar{Z}_j - z_{ij}}{\sigma_j} \tag{2.2}$$

Where,

$\bar{Z}_j$ = Sample mean for core component j

σ_j = Sample standard deviation for core component j

z_{ij} = Survey result of core component j for country i

Diverted and counterfeit drugs have a higher chance of entering the market when there is a lack of sufficient funding for healthcare which may result in shortage or erratic supply of drugs (63). This occurs more in developing countries which suffer from the lack of access to essential medicines (63). Total expenditure on healthcare as a percentage of Gross domestic product (GDP) by country was collected to measure the quality of healthcare in countries. According

to the World Bank, “GDP is the value of all final goods and services produced in a country in one year” (64). It is measured by adding an economy's incomes or expenditures and net exports.

Details related to the World Bank data on population, income level, unemployment percentage and GDP were used to evaluate the relationship between country attributes and the volume of diversion. These variables are of also of interest to counterfeiting. The number of unemployed people was calculated by multiplying unemployment percentage with the population of adults aged between 15 to 65 years.

According to WHO, developing countries have a higher percentage of counterfeit drugs due to lack of intellectual property right protection, regulatory and legal oversight, and presence of corruption in the health-care system (63). GDP and Country Income Level indicate if a country is a developing nation while Corruption Perception Index from Transparency International indicates the amount of corruption within a country. Similar to LPI, a higher value of Corruption Perception Index indicates lower corruption.

Data for country income level was obtained by collecting data for the adjusted net national income by the World Bank. Adjusted net national income is calculated by subtracting from gross national income a charge for the consumption of fixed capital and for the depletion of natural resources, which covers net forest depletion, energy depletion, and mineral depletion (64) . Results from Section 3.1 show that the country income level and GDP had a very high correlation. So, income per GDP was used in the regression and classification models to reduce multicollinearity. The formula used to obtain income per GDP is given below.

$$Income\ per\ GDP = \frac{Adjusted\ net\ national\ income}{GDP} \tag{2.3}$$

Transparency International, a global movement to end corruption calculated the Corruption Perception Index (CPI) in 2011 by using survey data from 17 sources. Results obtained from the survey were normalized and averaged for each country to obtain the CPI (65).

In the 'Transnational Organized Crime in East Asia and the Pacific: A Threat Assessment report' by UNODC (66), it is mentioned that counterfeiting of drugs is considered a violation of intellectual property rights rather than a violation of human rights in most of the countries when the drugs are patented. So, in order to establish better regulations, it is important for countries to have strong intellectual property rights which would make it harder for counterfeits to enter the legitimate supply chain. Information regarding regulations and enforcement of intellectual property rights was obtained through the Intellectual Property Rights Index (IPRI) developed by the World Intellectual Property Organization (WIPO). "WIPO is the global forum for intellectual property services, policy, information and cooperation" (67). IPRI is graded on a scale from 0 to 10 with 10 denoting the best intellectual property rights system.

Counterfeit drugs often go undetected and/or unreported. Prevalence of counterfeit drugs has not been made public mainly due to conflicts in the data available from different sources and since various organizations collecting this data maintain it as proprietary information (68). Pharmaceutical Security Institute maintains a Counterfeit Incident System (PSI CIS) database which contains the number of incidents reported in different regions. This data is collected from both open and non-public sources such as hospitals, warehouses, pharmacies, distributors, wholesalers. According to PSI, a "Counterfeit incident" is a report of a medicine deliberately and fraudulently mislabeled to make it look like a genuine generic/branded medicine (69). This data was used to classify the risk of counterfeiting into six levels.

Table 2-4 summarizes the variables of interest for Diversion and Counterfeiting along with the sources from which data was acquired.

Table 2-4 Variables of interest for Diversion and Counterfeiting

Variable	Data Source	Type of Variable	Range (raw-data)	Range (after pre-processing)
Previous Opioid abuse	UNODC	Independent, Continuous	0-2.5	0-2.5
Previous Opiate abuse	UNODC	Independent, Continuous	0.02-5.9	0.02-5.9
Previous Alcohol abuse	UNODC	Independent, Continuous	0.02-0.156%	3.2 -79,059,365.7
Population	World Bank	Independent, Continuous	35,000 – 1,337,705,000	35,000 – 1,337,705,000
Country Income Level (in million USD)	World Bank	Independent, Continuous	180,000,000- 1.26 e^{13}	180,000,000- 1.26 e^{13}
GDP (in million USD)	World Bank	Independent, Continuous	197 -15,000,000	197 -15,000,000
Women	World Bank	Independent, Continuous	0.25-0.55%	22,308-56,183,650
Unemployment	World Bank	Independent, Continuous	0-0.40%	181,214.6-649,128,660.7
Death Rate	WHO	Independent, Continuous	0-0.1015	0-0.1015
Logistics Performance Index (LPI)	World Bank	Independent, Continuous	1.34-4.09	1.34-4.09
Corruption perception index (CPI)	Transparency International	Independent, Continuous	0.9-9.4	0.9-9.4
Amount of money spent on healthcare (in million USD)	WHOSIS	Independent, Continuous	0.01-0.208%	1,030-19,541,041.6
Pharmaceutical imports (in million USD)	UN COMTRADE	Independent, Continuous	95,000-62,000,000,000	95,000-62,000,000,000
Cargo Theft Index (CTI)	FreightWatch International	Independent, Categorical	1-5	1-5
Intellectual Property Rights Index (IPRI)	WIPO	Independent, Continuous	2.9-8.6	2.9-8.6
Prescription drug seizure estimates as an estimate for diversion (in Kgs)	UNODC	Dependent, Continuous	0 – 23,000,000	0 – 23,000,000
Counterfeit incident Index (CII)	PSI	Dependent, Categorical	0-5	0-5

2.2 Regression analysis for predictive modeling

The data collected included around 180 observations of 17 variables. The goal of this study was to model the behaviors and incidence between these variables of interest and diversion

and counterfeiting. However, the data collected consisted of many missing values. Figures 2-1 to 2-4 show the pattern of missing values in the data collected for Diversion and Counterfeiting.

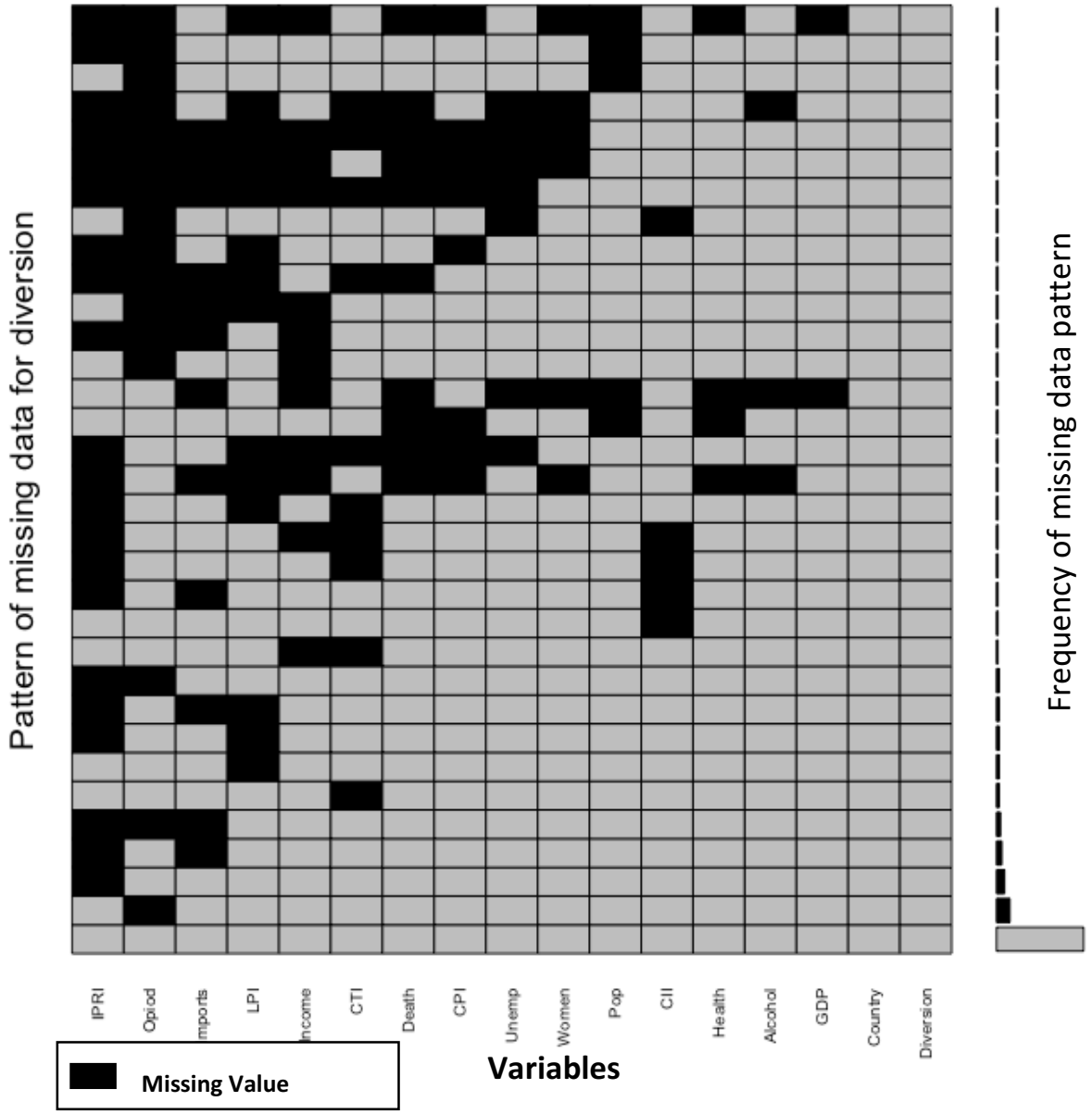

Figure 2-1Missing data pattern of variables used to construct model for diversion

Figure 2-1 suggests that the missing data pattern for variables used to construct a model for diversion is not completely random. This suggestion is also supported by IPRI missing values only for countries which have a higher Cargo Theft Index (CTI). Thus, it has been assumed

that the distribution of missingness in the data collected is Missing At Random (MAR), where the probability of a missing value depends on observed values of any variable in the dataset (70). Out of the 180 observations collected, 59 observations did not have data for diversion, so Figures 2-1 and 2-2 show the missing data pattern for the remaining 121 observations. There were only 80 complete cases among these 121 observations, which are shown in the last row of Figure 2-1.

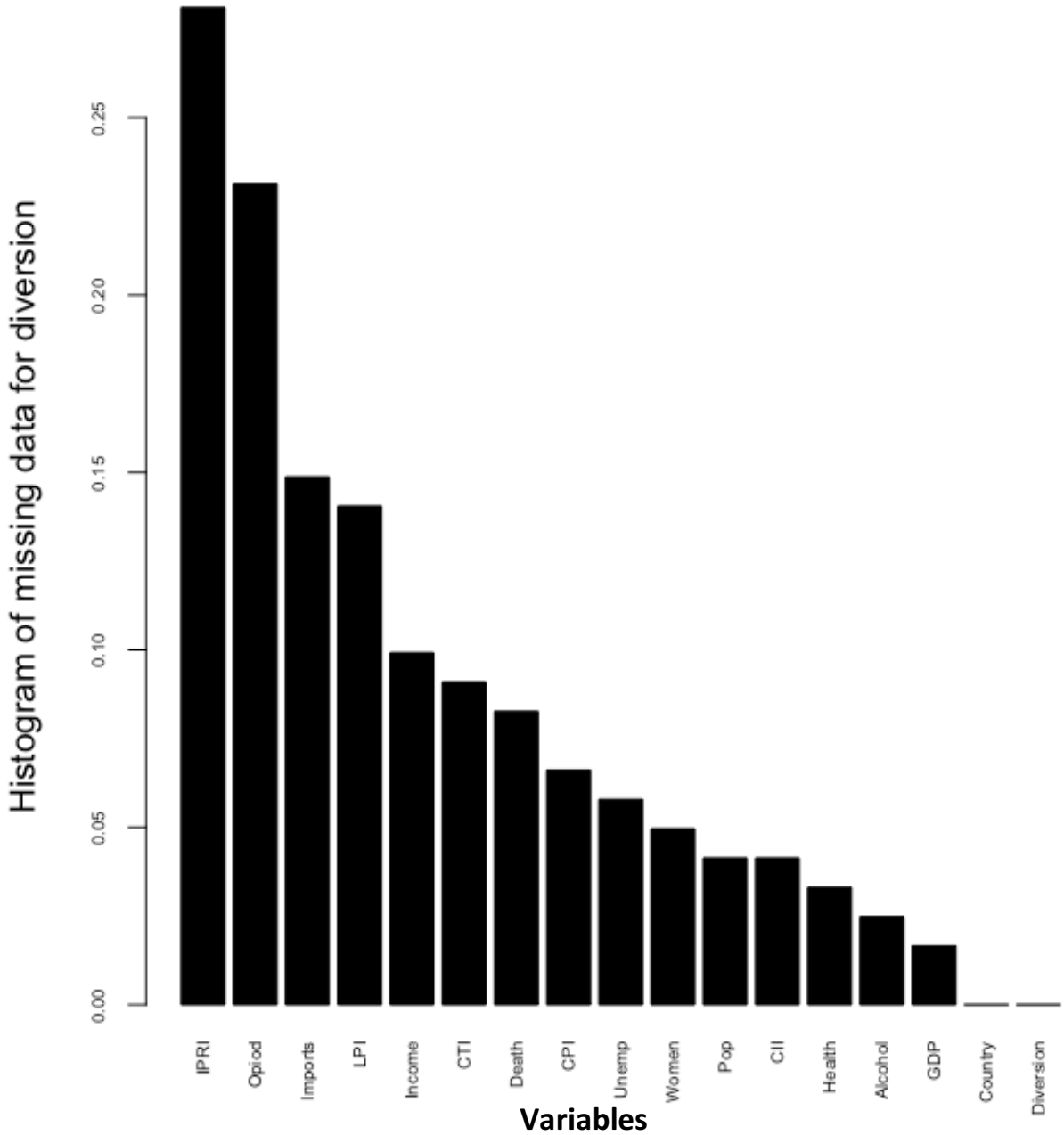

Figure 2-2 Histogram of missing values for variables used to construct model for diversion

IPRI had the highest missingness of 28.3% followed by Opioid use prevalence which had 23.1% missing, the rest of the variables had missingness over the range of 1% to 14.9%.

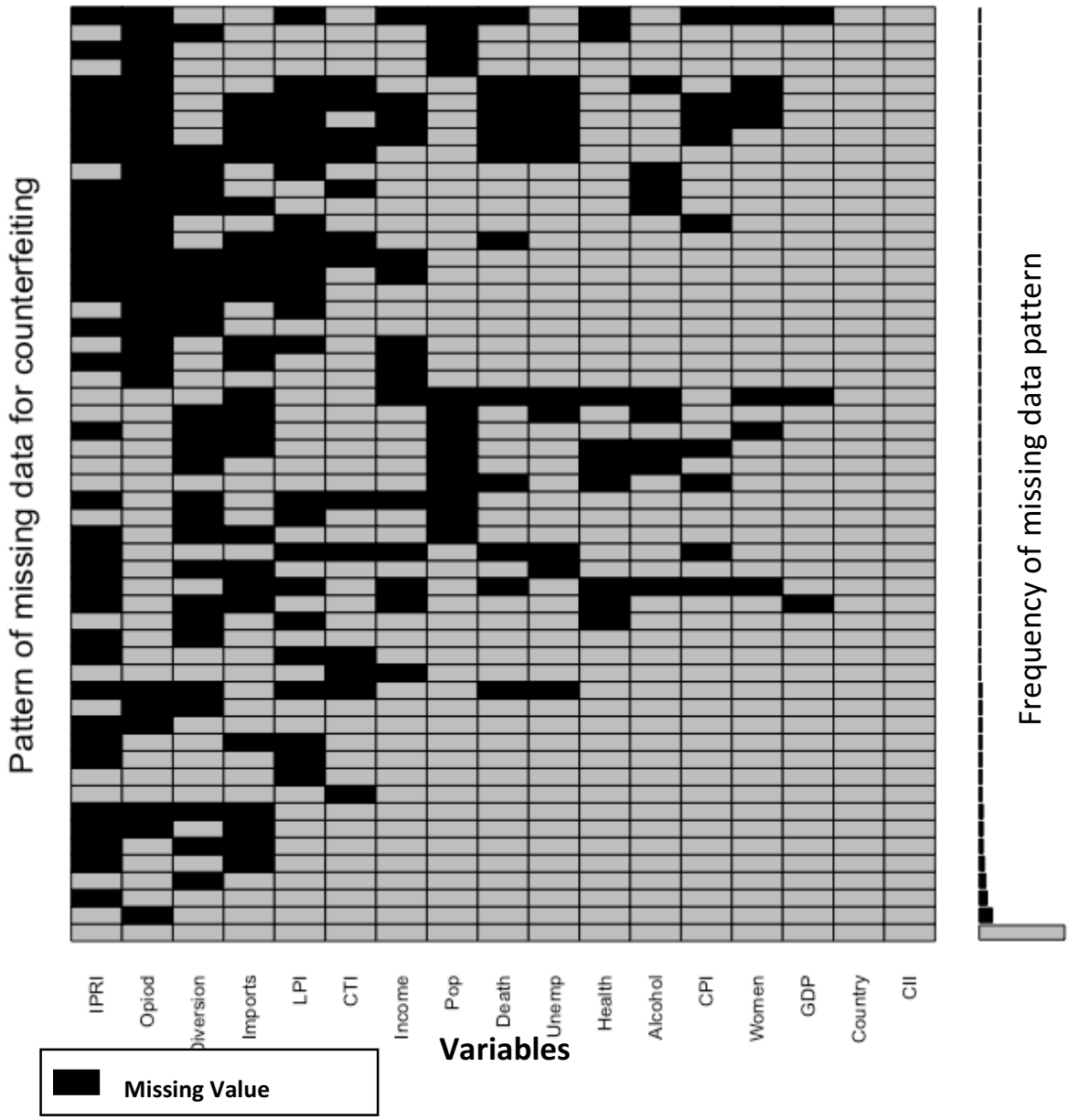

Figure 2-3 Missing data pattern of variables used to construct model for counterfeiting

The missing pattern shown in Figure 2-3 also suggests a MAR distribution for missing values. There are more number of observations here since only 21 out of the 180 observations collected did not have a value for Counterfeit Incidence index. However, the number of complete cases here was only 79.

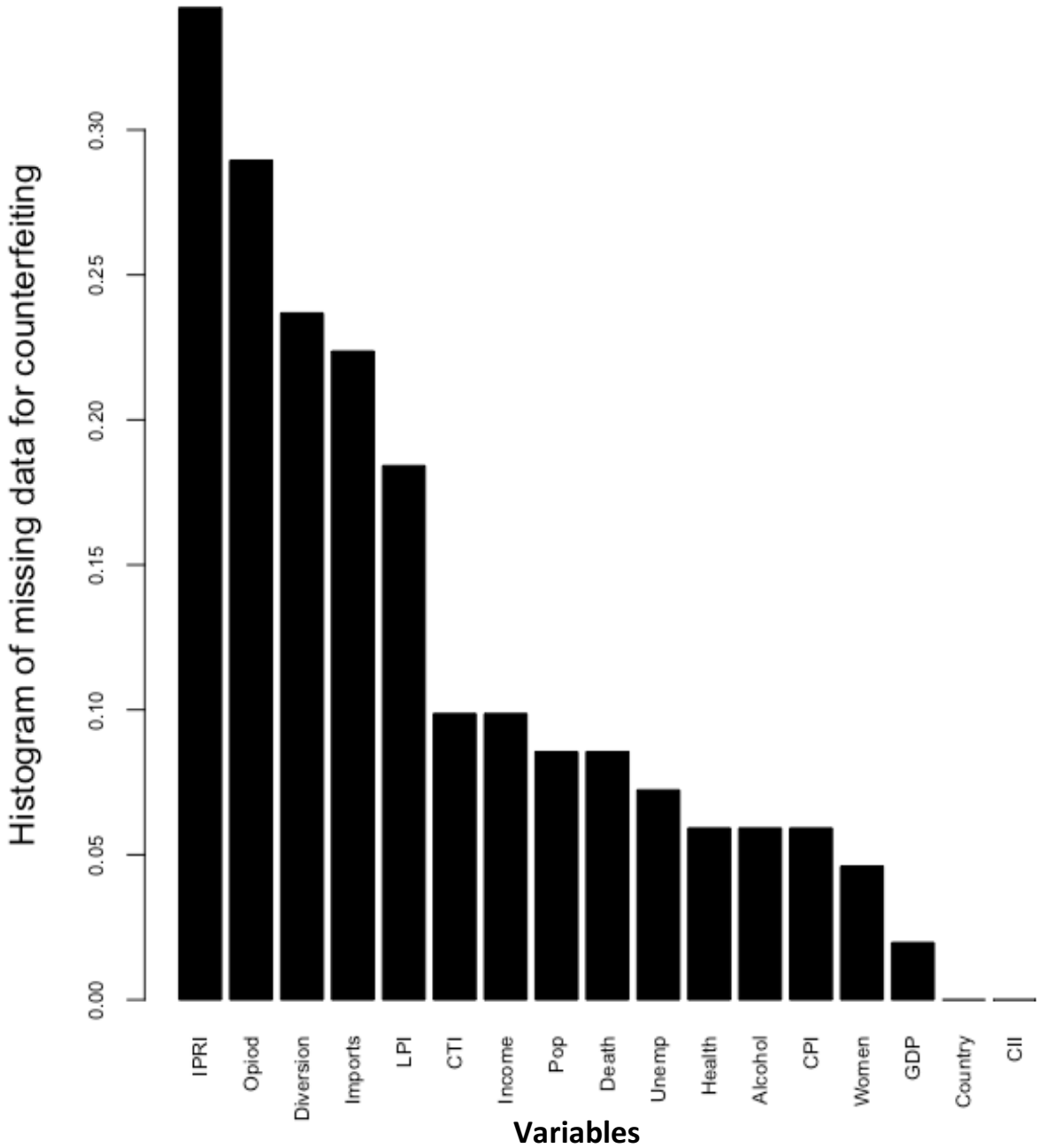

Figure 2-4 Histogram of missing values for variables used to construct model for counterfeiting

According to Figure 2-4, IPRI has the maximum number of missing values (34.2%), followed by opioid use prevalence (28.9%), diversion (23.7%), pharmaceutical imports (22.4%) and Logistics Performance Index (LPI) (18.4%).

The following sections explain how the collected data was analyzed to formulate a predictive model for estimating the volume of diversion and classifying regions into different levels of counterfeiting.

2.2.1 Diversion

There is no agreed upon statistic for measuring the illegal diversion of drugs during transportation, mainly due to cargo thefts being underreported (71). Absence of an international law which ensures consistency in reporting and tracking cargo thefts also adds to the issue of measuring incidents related to pharmaceutical cargo thefts (72). Due to this, prescription drug seizure estimates form WHO were considered to be an estimate of drugs diverted during transportation.

The data collected was analyzed in two ways; linear and non-linear regression. This was done to gain a better understanding of relationship between the diversion and the independent variables.

1. Linear Regression

Linear regression was used to answer the following questions:

- What factors influence diversion incidence?
- Is there a linear relationship between diversion and independent variables?
- Do the independent variables have an additive and/or interaction effects on the dependent variable?
- How much can the model explain with 80 observations?
- What is the predictive capability of a linear model?

Despite having fewer independent variables than the number of observations, a pairwise interaction model would increase the dimensionality of the model. Thus, in order to determine the relationship between the variables, and to discard variables which did not seem to have a linear relationship with diversion, the first step taken was to perform variable selection.

According to Fan & Lv (2008), Sure Screening is based on correlation learning which filters out the variables that have weak correlation with the response. "Sure Screening" refers to the property that all the important variables survive after variable screening with probability tending to 1" (73). The concept of sure independence screening is explained below (73) for our data:

If

p = number of predictors = 16

n = number of observations = 80

y = An $n \times 1$ response vector

X = A columnwise standardized $n \times p$ data matrix

β_i = Coefficient value for predictor i

Then,

Let $\mathcal{M}_* = \{for\ predictor\ i, 1 \leq \mathrm{i} \leq 16: \beta_i \neq 0\}$

be the true sparse model with no sparsity rate s = $|\mathcal{M}_*|$

Let $\omega = (\omega_1, \cdots, \omega_{17})^T$ be a vector obtained by the componentwise regression

i.e., $$\omega = \boldsymbol{X}^T \boldsymbol{y}$$

Then, for any given $\gamma \in (0,1)$, a submodel shrunk to size $\mathcal{M}_\gamma$ with a size of [80γ] < 80 can be computed by sorting the p componentwise magnitudes of the vector ω in a decreasing order,

i.e. $\mathcal{M}_\gamma = \{1 \leq \mathrm{i} \leq 16 : |\omega_i| \text{ is among the first } [80\gamma] \text{largest of all variables}\}$

This property of Sure Independence Screening (SIS) was used to limit the number of variables considering main as well as 2-way interaction effects to something less than the number of observations

The next step was to use variables selected by SIS to choose a linear regression model based on the explanatory power of the model as well as the mean squared error. A 5-fold cross

validation with replacement was used to reduce chances of over-fitting this model. Table 2.5 presents the number of observations in the set-up for model selection using 5-fold cross validation to determine the test and training datasets.

Table 2-5 Number of observations in Test and Train datasets using 5-fold Cross Validation

Fold		1	2	3	4	5
Number of observations in	Test	22	17	12	15	12
	Train	45	50	55	52	55

Model Selection involved using the following methods:

- **Ordinary Least Squares (OLS)**

This was the first method used to understand how the independent variables affected diversion from the data obtained. The main challenges in using this method were:

- Training datasets were small
- OLS has a higher chance of overfitting
- OLS does not perform variable selection
- OLS does not perform very well under multi-collinear conditions

The Matrix notation for OLS is given by:

$$\widehat{\boldsymbol{\beta}}^{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.4}$$

Where,

$\widehat{\boldsymbol{\beta}}^{OLS}$ = OLS estimate of coefficients

$\boldsymbol{X}$ = Design matrix

y = Vector of diverted drug seizure estimates

This model was developed in R using the linear model (lm) package. The code used is presented in the Appendix.

- **Ridge Regression**

Ridge regression is based on ordinary least squares. However, unlike OLS, this method introduces bias and lowers the variance of the model by increasing the value of a tuning parameter, λ. This involves using a tuning parameter(λ) to shrink the coefficient estimates of unimportant variables.

The matrix notation for Ridge regression is given by:

$$\widehat{\boldsymbol{\beta}}^R = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.5}$$

Where,

$\widehat{\boldsymbol{\beta}}^R$ = Ridge estimate of coefficients

$\boldsymbol{\lambda}$ = Ridge regression tuning parameter such that $\widehat{\boldsymbol{\beta}}^R = \begin{cases} \widehat{\boldsymbol{\beta}}^{OLS}, \text{when } \boldsymbol{\lambda} \to 0 \\ 0, \text{when } \boldsymbol{\lambda} \to \infty \end{cases}$

The ridge regression model was constructed in R using the glmnet package. The code is presented in the appendix.

- **Lasso Regression**

Least Absolute Shrinkage Selection Operator (Lasso) regression performs variable selection and regression analysis to find the model with a high predictive ability as well as better interpretability (74). This method is also based on Ordinary Least Squares and makes use of a tuning parameter like Ridge regression. However, unlike Ridge regression, Lasso shrinks the coefficients of unimportant variables as well as highly correlated independent variables to zero, thus dropping them and completing the variable selection. The Lasso coefficient estimate is obtained by:

$$\widehat{\boldsymbol{\beta}^L} = arg\min_{\beta\in\mathbb{R}^p}\left\{\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{2.6}$$

Where,

$\widehat{\boldsymbol{\beta}}^L$ = Lasso estimate of coefficients

$\boldsymbol{\lambda}$ = Lasso regression tuning parameter ≥ 0

p = number of variables

n = number of observations

After obtaining the coefficient values for the training datasets through the abovementioned methods, the test datasets were used to test the predictive ability through:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}^{\boldsymbol{test}}\hat{\boldsymbol{\beta}} \tag{2.7}$$

And the mean squared error was calculated using:

$$\text{Mean squared error} = \frac{1}{n}\ (\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}}) \tag{2.8}$$

Where,

n = number of observations in the test set

$\boldsymbol{X}^{\boldsymbol{test}}$ = Test Data set

The adjusted R^2 was calculated using:

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - p - 1}\right] \tag{2.9}$$

The Lasso model was constructed in R using the glmnet package, the code is presented in the appendix.

Figure 2-5 summarizes steps involved in the linear regression based analysis.

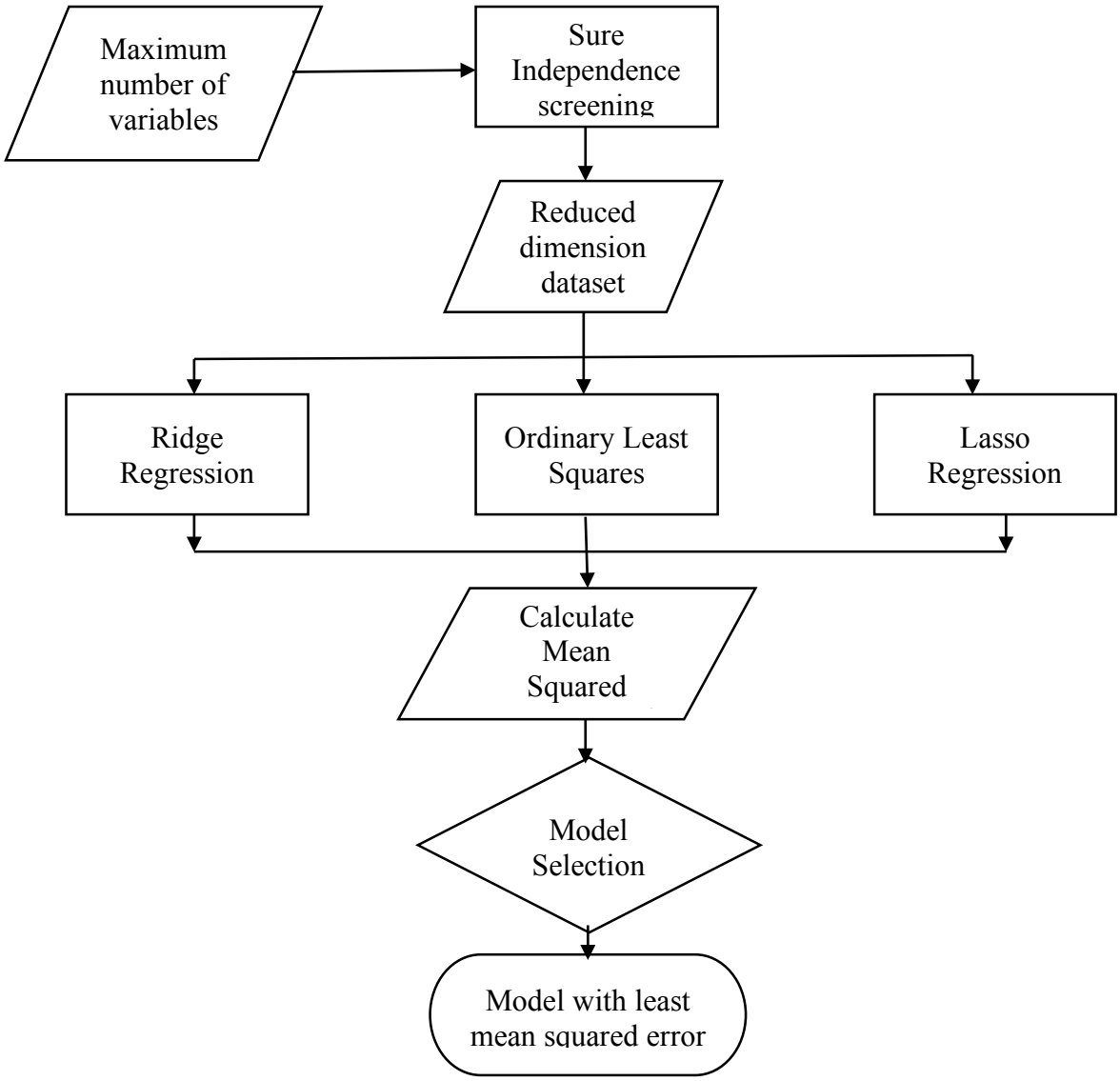

Figure 2-5 Linear Regression based Analysis

The results of these linear regression models are discussed in Chapter 3 in Section 3.1

2. *Nonlinear Regression*

Linear models are useful for understanding additive and linear effects. But they do not perform as well when the relationship between predictors and the response is complex or nonlinear. Linear regression methods also run into problems when they encounter missing values;

because of which they exclude rows having missing values, thus reducing the size of the dataset.

Nonlinear regression methods are useful in these scenarios. Methods such as regression trees represent nonlinear relationships and include rows having missing values in the model. Regression trees are based on a recursive partitioning technique. According to Strobl, Malley, & Tutz, 2009, in regression trees, the space containing predictor variables is partitioned into rectangular areas such that all predictor variables having similar response values are grouped together. After partitioning, a constant value of the response variable is predicted for each area, thus forming a piecewise constant model (75). Accuracy is measured by obtaining the mean squared error within each area. This process of splitting is repeated until the specified stopping criteria is satisfied. These qualities provided motivation for using regression trees to analyze the data. Number of observations included in this step increased from 80 to 121. This data was randomly divided into 80% training and 20% test data sets.

The splitting criteria used for recursive partitioning was to minimize impurity of a node, which is the mean squared error within a node. The stopping criteria was to stop splitting when the mean squared error improved by less than 0.001.

Strobl et al., (2009), used ensemble methods such as bagging and random forests instead of just a single regression tree. These methods are used to improve the model by making it robust to withstand higher variability in the data (75). These methods base their prediction on a larger number of regression trees instead of just one tree. Each of these trees are built using bootstrapped samples, i.e., samples randomly drawn with replacement from the training dataset.

Bagging, also known as bootstrap aggregation, “is a method for generating multiple versions of a predictor and using these to get an aggregated predictor” (76). Bagging has been shown to improve prediction accuracy.

Random forests make use of bootstrapped samples while randomly restricting the set of predictors in each split. This makes the trees more diverse and allows predictors having weaker effects to enter the ensemble. Breiman (2001) concluded that random forests in general, do not overfit due to the law of large numbers. They also work towards reducing bias and variance without progressively changing the training set (77).

Figure 2-6 shows the outline of construction of random forests given by Breiman (2002) adopted to construct the random forest in this study (78). The model for this study was constructed in R using the randomForest package.

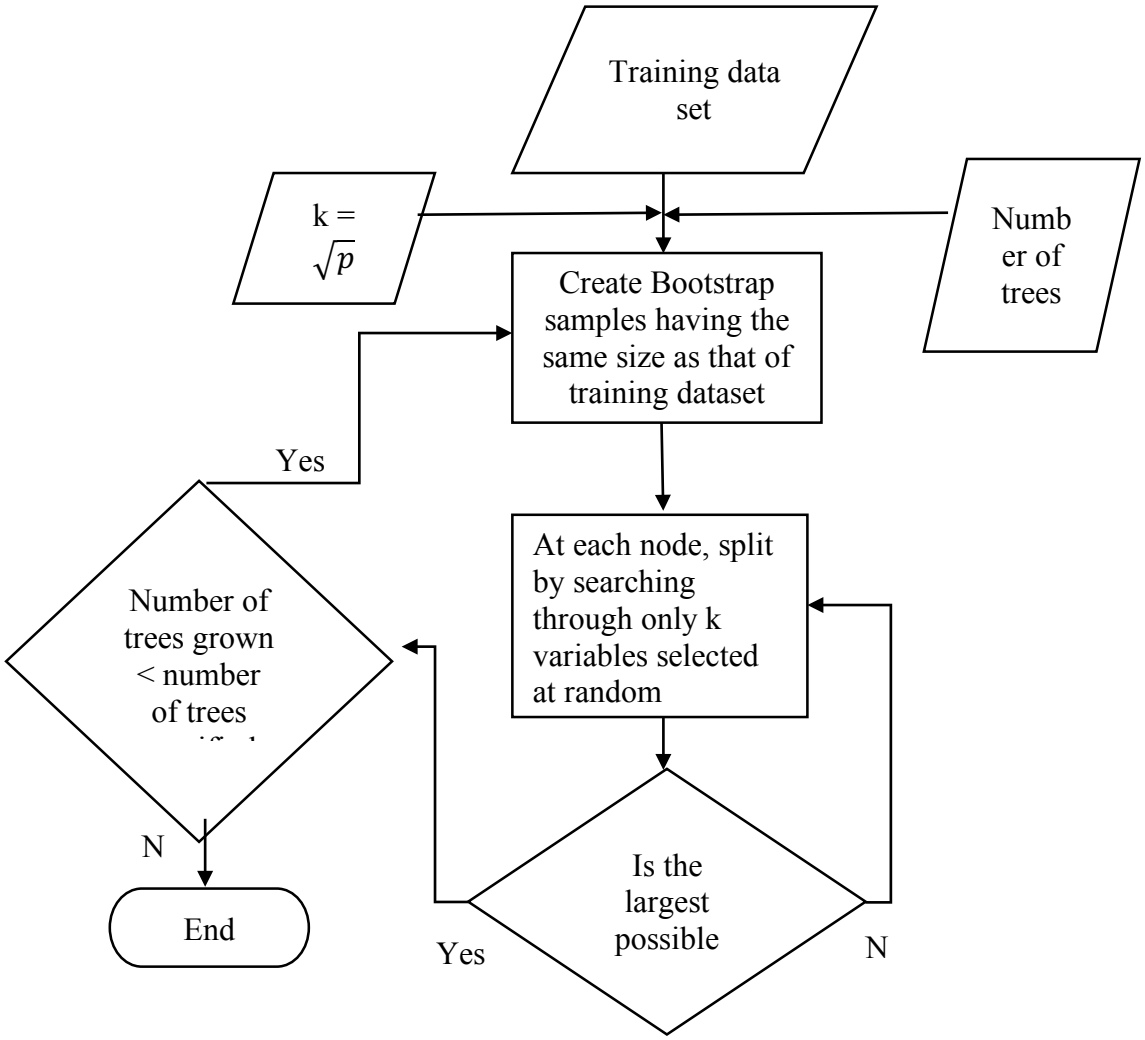

Figure 2-6 Construction of Random Forest

Where,

k = Number of variables in each split selected at random

p = Number of predictors

The initial value of k was taken as the square root of the number of predictors and the number of trees to grow was set at 2000. These two parameters were varied to find their best value which would reduce the mean squared error and increase the variance explained by the model. These results are presented in Chapter 3 in Section 3.2.

2.2.2 Counterfeiting

The data collected for counterfeiting of drugs assigns an estimated counterfeit incident index to each country based on the number of incidents reported in the region. This introduces a lot of uncertainty since this is not an actual indicator of the number of incidents in a region. However, due to the lack of accurate and complete information regarding these incidents, the collected data was considered to be an ordered, categorical variable representing the levels of counterfeiting. Table 2-6 shows the classification of counterfeit incident index based on the number of incidents.

Table 2-6 Classification of counterfeit incident index

Number of Counterfeiting incidents	Counterfeit incident index
0	0
1-5	1
6-25	2
26-100	3
101-200	4
>201	5

The relationship between counterfeiting and all the independent variables mentioned in Table 2-4 was assessed using Ordinal logistic regression by constructing an ordinal logistic regression model using the neural nets (nnets) package in R. Both additive and interaction effects were considered and the model which had a higher classification accuracy was selected for the predictive model.

Ordinal logistic regression is one of the most commonly used classification method for ordered categories. It is the extension of logistic regression which is used for dichotomous responses. The proportional odds model is a type of ordinal logistic regression which simultaneously considers the effects of predictors across all levels of the ordered response variable (79). This model is based on the parallel slopes assumption. According to this assumption, the dependent variable's levels are parallel to each other because the correlation between the dependent and the independent variables is the same for all levels of the dependent variable (80).

The proportional odds model is given by (81):

$$P(Y \geq j|X) = \frac{1}{1 + e^{-(\alpha_j + \boldsymbol{X\beta})}} \tag{2.10}$$

Where,

j = number of ordered levels = 6

α_j= intercepts

β = Coefficients

The predicted values from this model give the probability of the response Y being in level j and above versus all levels below j.

All variables from Table 2-4, except the counterfeiting incident index were assumed to be independent variables. The dataset consisted of 79 observations, 16 independent and 1 dependent variable which was randomly divided into 80% training and 20% test datasets.

2.2.3 Prediction

After obtaining the best predictive model with the lowest mean squared error for diversion and the lowest classification error for counterfeiting, a new dataset was created to implement these models. This dataset contains data from countries belonging to an existing distribution network in South-East Asia (82). Agility Global Integrated Logistics, a global logistics provider based out of Kuwait with a network across 100 countries developed an integrated trucking network in Southeast Asia (82). They are also one of the global leaders in pharmaceutical logistics. Figure 2-7 describes their integrated trucking distribution network in Southeast Asia.

Information was collected for the countries present in this distribution network as shown in Tables 2-7 and 2-8. This information was obtained from the same sources mentioned in Table 2-4, however, this information was obtained for the year 2015.

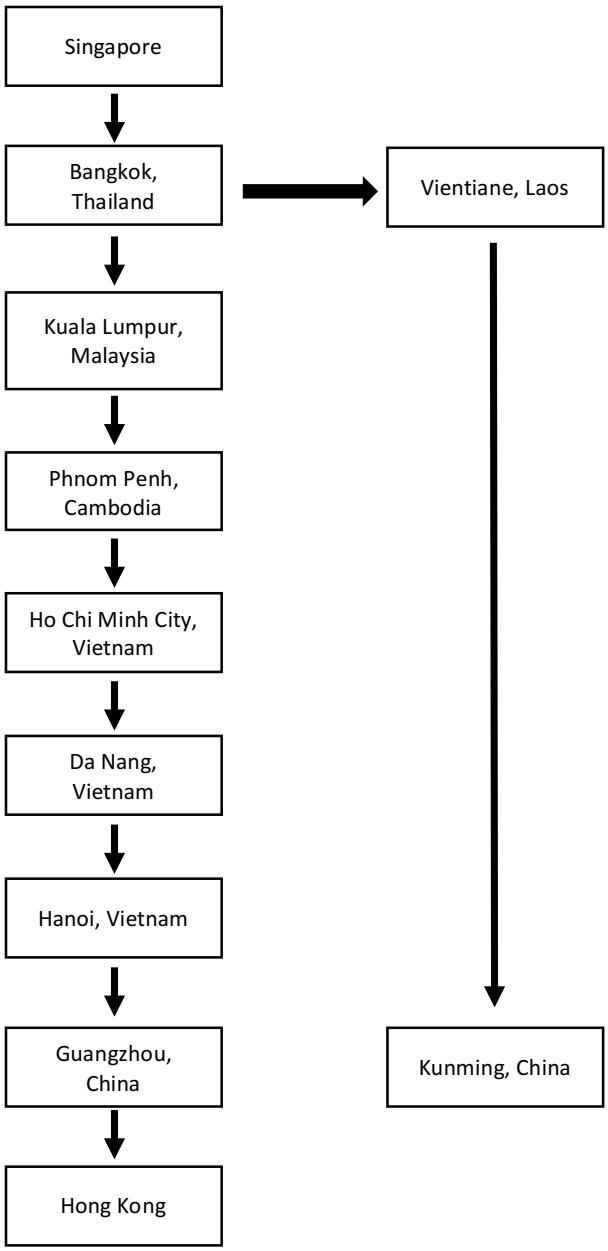

Figure 2-7 Agility's road freight transportation network in Southeast Asia (Source: http://www.agility.com/)

Table 2-7 Data for countries from Agile logistics' pharmaceutical distribution network

Country	Total Population	% Unemployment	% Women	GDP	% GDP on health	Alcohol Consumption	CPI	Death Rate	LPI	Cargo theft index	Country income level	IPRI
Malaysia	30,331,007.00	2.90	48.40	2.96218E+11	4.17	0.82	50	4.92	3.59	3	2.50571E+11	6.8
Cambodia	15,577,899.00	0.30	51.22	18,049,954,289	5.68	4.77	21	6.13	2.74	4	13,646,958,405	7.9
Thailand	67,959,359.00	0.80	50.71	3.95282E+11	6.53	7.08	38	7.90	3.43	2	3.03389E+11	5
Vietnam	91,703,800.00	1.80	50.53	1.93599E+11	7.07	3.77	31	5.82	3.15	2	1.40419E+11	4.7
Laos	6,802,023.00	1.40	50.23	12,327,488,341	1.87	6.73	25	6.77	2.39	2	8,757,477,492	5.2
China	1,371,220,000.00	4.10	48.48	1.08664E+13	5.55	5.91	37	7.20	3.53	3	8.7579E+12	5.4
Hong Kong	7,305,700.00	3.20	53.05	3.09929E+11	5.40	2.83	75	6.20	3.83	3	1.98853E+11	7.8

Tables 2-8 to 2-11 show the distance matrix for the existing route used by Agility (Figure 2-7) and the distance matrices used for different modes of transportation to calculate the costs mentioned in Section 2.3. All the distances are in kilometers.

Table 2-8 Distance matrix for the route currently used by Agility

	Singapore	Kuala Lumpur	Bangkok	Vientiane	Phnom Penh	Ho Chi Minh City	Da Nang	Hanoi	Guangzhou	Kunming	Hong Kong
Singapore		355									
Kuala Lumpur			1472								
Bangkok				653	859						
Vientiane										850	
Phnom Penh						279					
Ho Chi Minh City							900				
Da Nang								780			
Hanoi									799		
Guangzhou											120
Kunming											
Hong Kong											

Table 2-9 Distance matrix when the mode of transportation is air

	Singapore	Kuala Lumpur	Bangkok	Vientiane	Phnom Penh	Ho Chi Minh City	Da Nang	Hanoi	Guangzhou	Kunming	Hong Kong
Singapore		327.01	1,474.02	1,872.48	1,152.22	1,111.65	1,709.15	2,253.25	2,634.61	2,663.57	2,606.44
Kuala Lumpur	327.01		1,237.27	1,671.26	1,011.39	1,029.56	1,606.81	2,095.00	2,558.92	2,464.14	2,542.98
Bangkok	1,474.02	1,237.27		536.59	563.53	776.42	878.60	1,037.42	1,712.17	1,306.03	1,742.94
Vientiane	1,872.48	1,671.26	536.59		776.35	931.14	650.31	535.62	1,268.25	818.27	1,319.86
Phnom Penh	1,152.22	1,011.39	563.53	776.35		230.10	627.29	1,117.98	1,579.07	1,546.92	1,574.19
Ho Chi Minh City	1,111.65	1,029.56	776.42	931.14	230.10		615.71	1,198.74	1,551.74	1,665.79	1,532.31
Da Nang	1,709.15	1,606.81	878.60	650.31	627.29	615.71		659.91	956.34	1,175.12	951.59
Hanoi	2,253.25	2,095.00	1,037.42	535.62	1,117.98	1,198.74	659.91		831.01	580.41	905.38
Guangzhou	2,634.61	2,558.92	1,712.17	1,268.25	1,579.07	1,551.74	956.34	831.01		1,121.91	147.17
Kunming	2,663.57	2,464.14	1,306.03	818.27	1,546.92	1,665.79	1,175.12	580.41	1,121.91		1,516.97
Hong Kong	2,606.44	2,542.98	1,742.94	1,319.86	1,574.19	1,532.31	951.59	905.38	147.17	1,516.97	

Table 2-10 Distance matrix when the mode of transportation is road

	Singapore	Kuala Lumpur	Bangkok	Vientiane	Phnom Penh	Ho Chi Minh City	Da Nang	Hanoi	Guangzhou	Kunming	Hong Kong
Singapore		354.76									
Kuala Lumpur			1,471.93								
Bangkok		1,471.93		646.47	659.38						
Vientiane			646.47		973.99	1,253.71	895.18	749.19	1,247	850	
Phnom Penh			659.38	973.99		279.79	971.75	1,493.46			
Ho Chi Minh City				1,253.71	279.79		858.64	1,619.09	1,540	1,615	
Da Nang				895.18	971.75	858.64		767.76	950	1,130	
Hanoi				749.19	1,493.46	1,619.09	767.76		801.24	528	
Guangzhou				1,247		1,540	950	801.24		1,372.37	119
Kunming				850		1,615	1,130	528	1,372.37		1,192
Hong Kong									119	1,192	

Table 2-11 Distance matrix when the mode of transportation is water

	Singapore	Kuala Lumpur	Bangkok	Vientiane	Phnom Penh	Ho Chi Minh City	Da Nang	Hanoi	Guangzhou	Kunming	Hong Kong
Singapore		446.06	1,554.47		1,369.32	1,220.76	3,363.74	2,656.43	2,805.71		2,744.94
Kuala Lumpur	446.06		1,952.13		1,767.43	1,618.86	3,761.85	3,054.54	3,203.82		3,143.04
Bangkok	1,554.47	1,952.13			1,150.54	1,232.47	3,389.11	2,681.80	2,891.76		2,774.21
Vientiane											
Phnom Penh	1,369.32	1,767.43	1,150.54			941.87	3,098.51	2,391.20	2,575.63		2,483.61
Ho Chi Minh City	1,220.76	1,618.86	1,232.47		941.87		2,374.56	1,667.25	1,851.68		1,759.66
Da Nang	3,363.74	3,761.85	3,389.11		3,098.51	2,374.56		956.88	1,793.74		1,715.68
Hanoi	2,656.43	3,054.54	2,681.80		2,391.20	1,667.25	956.88		1,084.85		1,006.79
Guangzhou	2,805.71	3,203.82	2,891.76		2,575.63	1,851.68	1,793.74	1,084.85			182.52
Kunming											
Hong Kong	2,744.94	3,143.04	2,774.21		2,483.61	1,759.66	1,715.68	1,006.79	182.52		

2.3 Cost Functions

The cost of one unit of a drug generally varies from country to country. Table 2-12 shows the cost of 1mg of alprazolam, a sedative used to treat anxiety and panic disorders across different countries in South-east Asia. This data was collected from Drugs.com, an independent medicine information website that provides a free drug-information service (83).

Table 2-12 Cost of one unit of Alprazolam by country

Country	Cost of 1mg of Alprazolam (in US Dollars)
Malaysia	0.51
Cambodia	0.40
Thailand	0.51
Vietnam	0.38
Laos	0.38
China	0.45
Hong Kong	0.51

This information is used as described in Sections 2.3.1 – 2.3.4 for estimating the costs involved.

2.3.1 Cost of Diversion

The Chubb Group of Insurance Companies estimates that pharmaceutical companies usually lose $4.00 to non-reimbursable indirect costs for every $1.00 that they recover from insurance companies after experiencing a cargo theft (84). These are attributed to the following:

a. Lost Sales
b. Replacement Shipments
c. Customer Dissatisfaction
d. Implementation of theft deterrence

FreightWatch International's Supply Chain Intelligence Center (SCIC) reported that 75% of cargo thefts in Europe, Middle East and Africa (EMEA) occurred while freight was being transported on road. Railways and waterways in the EMEA experienced 10% of cargo thefts

while the airways experienced 5% (85). Road transportation has the highest risk of facing pharmaceutical cargo theft, followed by railways, waterways and airways.

For a drug X let,

C = Cost of one-unit of drug in the given region in USD

$\widehat{D}$ = Estimated amount of diversion from the predictive model in Kg

m = mode of transport = {road, rail, water, air}

$$M(m) = \begin{cases} 1, \text{if m} = \text{mode of transport used} \\ 0, \text{if m} \neq \text{mode of transport used} \end{cases}$$

$$P(m) = \text{Percentage of diversion based on mode of transport} = \begin{cases} 0.75, \ m = road \\ 0.1, \ m = rail \\ 0.1, \ m = water \\ 0.05, \ m = air \end{cases}$$

The expected cost of diversion is then given by:

$$E[\text{Cost of Diversion for an arc}] = 4 \times C \times P(m) \times \widehat{D} \tag{2.11}$$

2.3.2 Cost of Counterfeiting

According to an analysis by WHO in 2000, 1% of counterfeits were copies of the original product, 15.6% had the right ingredients but fake packaging, 32.1% lack active ingredients, 20.2% had incorrect quantities of active ingredients, 21.4% contained wrong ingredients, 8.5% contained high levels of impurities (86).

Considering the effect of counterfeit drugs on a manufacturer, when a counterfeit drug is the copy of an original or has the right ingredients but fake packaging and goes undetected, the manufacturer suffers the loss of that unit of drug. If the counterfeit copy does not work due to lack or incorrect quantities of active ingredient, the manufacturer faces loss from the counterfeit copy as well as future business from that particular patient.

The Counterfeit Incident Index level predicted is not a direct indicator of the number of counterfeited units. Data regarding the volume of counterfeiting is usually not reported on the

same scale, which makes it difficult to store this information in a way that is accepted by everyone (87). There is some evidence in the literature that this might vary from a few hundreds to a few thousands (87). This was used as the basis for obtaining cost of counterfeiting by assuming volume of counterfeiting to be uniformly distributed between 100 to 999,999 units per incident.

Counterfeit drugs having impurities or wrong ingredients pose a safety risk to the health of a patient (87). An estimated 700,000 Africans die every year from counterfeit anti-malarial and tuberculosis drugs which causes a loss of $12 billion USD in economic output to the African nations (88), which implies that an average of $17142.85 USD is lost per person. Assuming that each individual buys an average of 10 units, the sale of one harmful counterfeit unit results in the loss of $1714.85 USD.

For a drug X let,
C = Cost of one-unit of drug in the given region in USD
S = Estimated Counterfeit Incident Index
n = number of incidents per estimated Counterfeit Incident Index
q = Average number of counterfeit copies detected per incident reported ~ U (100,999999)

Percentage of counterfeit copies which work the same as the original = 16.6%
Percentage of counterfeit copies which do not work = 52.3%
Percentage of counterfeit drugs which might be harmful = 29.9%

Cost of counterfeit copies which work the same as the original = 0.166C
Cost of counterfeit copies which do not work = 2C × 0.523 = 1.046C
Cost of counterfeit copies which might be harmful = (1714.285) × 0.299

$$\text{E [Cost of counterfeiting for an arc]} = \mathrm{S} \times \mathrm{n} \times \mathrm{q} \times \left(\sum_{Type\ of\ Copy} Cost\ of\ counterfeit\ copies \right) \tag{2.12}$$

2.3.3 Cost of Transportation

The cost of transportation depends on both distance covered during transportation and the mode of transportation.

Let,
d(m) = Distance of transport as a function of mode of transport
T(m) = Cost of transportation per unit distance as a function of mode of transport

Then,

$$\text{Cost of transportation for an arc} = d(m) \times T(m) \tag{2.13}$$

Table 2-13 shows the cost of transportation per unit distance for different modes of transportation. These costs were obtained by using the freight calculator from World Freight Rates (89). All costs were calculated with Singapore as the origin and for transporting 5 tons of pharmaceutical products.

Table 2-13 Transportation cost per unit distance

Cost per unit distance (in US Dollars)			
Road (per km)	**Rail (per km)**	**Water (per km)**	**Air (per km)**
2.44	1.44	1.68	12.2

2.3.4 Expected total cost

Expected total cost of transporting drugs across each arc, i.e., a direct connection between any two adjacent countries can be computed by adding the cost of diversion, counterfeiting and transportation. This is given by:

$$
\begin{aligned}
&Expected\ total\ cost\ for\ an\ arc \\
&\quad = 4 \times \mathrm{C} \times \mathrm{P(m)} \times \widehat{D} \\
&\quad + \ (\mathrm{S} \times \mathrm{q} \times \mathrm{n}) \left(\sum_{Type\ of\ Copy} Cost\ of\ counterfeit\ copies \right) \\
&\quad + \ \big(d(m) \times T(m)\big)
\end{aligned}
\tag{2.14}
$$

2.4 Estimation of best route

The primary goal of this study was to construct a model to estimate optimal routes for transporting drugs, which is essentially an optimization problem. This problem can be visualized as a network which needs to be optimized by reducing all the costs involved. In other words, this can be constructed as a shortest path problem whose objective is to find the least expensive path for transporting drugs.

Dijkstra's algorithm is a single source shortest path algorithm which can be used to find the least expensive path from a source to different destinations (90). Using this algorithm, the problem of finding the shortest path can be visualized as a weighted graph. The vertices of the graph represent countries; their corresponding arc weights represent the cost associated with transporting drugs between those countries. The objective is to find the least expensive route and can be represented by:

$$
\begin{aligned}
\boldsymbol{Objective}: \min_{\text{routes}} \Bigg[\sum_{\text{arcs}} \Bigg\{ & 4 \times \mathrm{C} \times \mathrm{P(m)} \times \widehat{D} \\
& + \ (\mathrm{S} \times \mathrm{q} \times \mathrm{n}) \left(\sum_{Type\ of\ Copy} Cost\ of\ counterfeit\ copies \right) \\
& + \ \big(d(m) \times T(m)\big) \Bigg\} \Bigg]
\end{aligned}
$$

Results obtained from Equation 2.11 were used to create the distance matrix that was then used to form the weighted graph. The algorithm used can be described as given below (90):

Let G = (V, A) weighted graph representing the problem,

Where,

V(G) = $\{v_0, v_1, \ldots, v_k\}$ Set of vertices representing countries

A(G) = $\{a_0, a_1, \ldots, a_i\}$ Set of all arcs connecting adjacent vertices

$w(uv) \geq 0$ represents the cost of entering country v from country u

O = Set of countries to which the minimum cost route from origin u=v_0 is known

Q = {v, v ≠ u}, Set of countries in queue for which the minimum cost route from origin u is unknown, such that, O ∩ Q = Ø

t (v) = Cost of the least expensive route found yet from origin u to country v

Input: G = (V, A)

Initialization:

i. O = {u}, cost from u to u = 0
ii. t (v) = $w(uv)$ for all v ≠ u

Iteration:

i. Select country v_k from Q such that t (v_k) is the minimum in the set {t (v) | v ∉ O}
ii. Add v_k to O
iii. For each arc (v_k, v_{k+1}) from country v_k to country v_{k+1} ∉ O,
update t (v_{k+1}) $= \min\ \{t\ (v_{k+1}), \mathrm{w}(\mathrm{u}, v_k) + w(v_k, v_{k+1})\}$

Termination: Continue iteration until,

O = V(G) ⇒ All minimum cost routes from u have been found

or

t (v_k) = ∞ for every v_k ∉ O ⇒ Remaining countries are unreachable from u

Expected costs calculated vary across different routes and modes of transportation. With the presence of 10 nodes, there are 3,628,800 (i.e., 10!) ways in which these 10 nodes can be positioned along different routes. Along with these combinations of routes, the presence of

different modes of transport adds to the complexity of the graph by adding parallel edges to the graph.

In the presence of these complexities, the computational time increases for Dijkstra's algorithm. To reduce this complexity and the computational time, parallel arcs which are more expensive were eliminated to form a simple graph as the input for Dijkstra's algorithm.

3 Results

3.1 Diversion

One of the goals of this thesis was to understand and define the relationship between diversion, counterfeiting and other variables that can affect the risk of diversion and counterfeiting. The first steps taken here was to obtain the correlation matrix. Table 3.1 shows the correlation matrix between all the dependent and independent variables.

Table 3-1 Correlation matrix for dependent and independent variables

	Opioid	Pop	Unemployment	Women	GDP	Health	Alcohol	CPI	Death	Diversion	LPI	CII	CTI	Income	Imports	IPRI	Income per GDP
Opioid	1.00	0.14	0.08	0.12	0.76	0.39	0.22	0.17	0.51	0.00	0.18	0.07	0.21	0.74	0.66	0.22	-0.03
Pop	0.14	1.00	-0.16	-0.05	0.53	-0.12	-0.02	-0.10	-0.07	0.11	0.11	0.67	0.11	0.51	0.21	-0.06	-0.04
Unemployment	0.08	-0.16	1.00	0.23	-0.03	0.20	0.29	-0.09	0.31	-0.02	-0.21	-0.20	-0.08	0.08	-0.01	-0.14	0.52
Women	0.12	-0.05	0.23	1.00	0.00	0.27	0.36	-0.08	0.25	0.00	-0.13	-0.08	0.03	0.00	0.06	-0.10	0.00
GDP	0.76	0.53	-0.03	0.00	1.00	0.34	0.13	0.19	0.27	0.07	0.33	0.23	0.20	0.98	0.81	0.25	-0.04
Health	0.39	-0.12	0.20	0.27	0.34	1.00	0.51	0.54	0.31	-0.18	0.48	-0.26	0.09	0.37	0.51	0.51	0.17
Alcohol	0.22	-0.02	0.29	0.36	0.13	0.51	1.00	0.21	0.42	-0.22	0.25	-0.03	0.09	0.14	0.25	0.23	0.09
CPI	0.17	-0.10	-0.09	-0.08	0.19	0.54	0.21	1.00	0.25	-0.12	0.84	-0.41	-0.23	0.18	0.36	0.94	-0.07
Death	0.51	-0.07	0.31	0.25	0.27	0.31	0.42	0.25	1.00	-0.12	0.10	-0.21	-0.15	0.25	0.27	0.24	-0.07
Diversion	0.00	0.11	-0.02	0.00	0.07	-0.18	-0.22	-0.12	-0.12	1.00	-0.08	0.14	0.02	0.07	0.01	-0.14	-0.01
LPI	0.18	0.11	-0.21	-0.13	0.33	0.48	0.25	0.84	0.10	-0.08	1.00	-0.12	-0.07	0.31	0.47	0.88	-0.09
CII	0.07	0.67	-0.20	-0.08	0.23	-0.26	-0.03	-0.41	-0.21	0.14	-0.12	1.00	0.40	0.23	0.09	-0.34	0.01
CTI	0.21	0.11	-0.08	0.03	0.20	0.09	0.09	-0.23	-0.15	0.02	-0.07	0.40	1.00	0.19	0.25	-0.18	-0.04
Income	0.74	0.51	0.08	0.00	0.98	0.37	0.14	0.18	0.25	0.07	0.31	0.23	0.19	1.00	0.79	0.22	0.17
Imports	0.66	0.21	-0.01	0.06	0.81	0.51	0.25	0.36	0.27	0.01	0.47	0.09	0.25	0.79	1.00	0.41	-0.05
IPRI	0.22	-0.06	-0.14	-0.10	0.25	0.51	0.23	0.94	0.24	-0.14	0.88	-0.34	-0.18	0.22	0.41	1.00	-0.14
Income per GDP	-0.03	-0.04	0.52	0.00	-0.04	0.17	0.09	-0.07	-0.07	-0.01	-0.09	0.01	-0.04	0.17	-0.05	-0.14	1.00

3.1.1 Linear Regression

Linear regression was used to examine additive and interaction effects of the independent variables on illegal diversion.

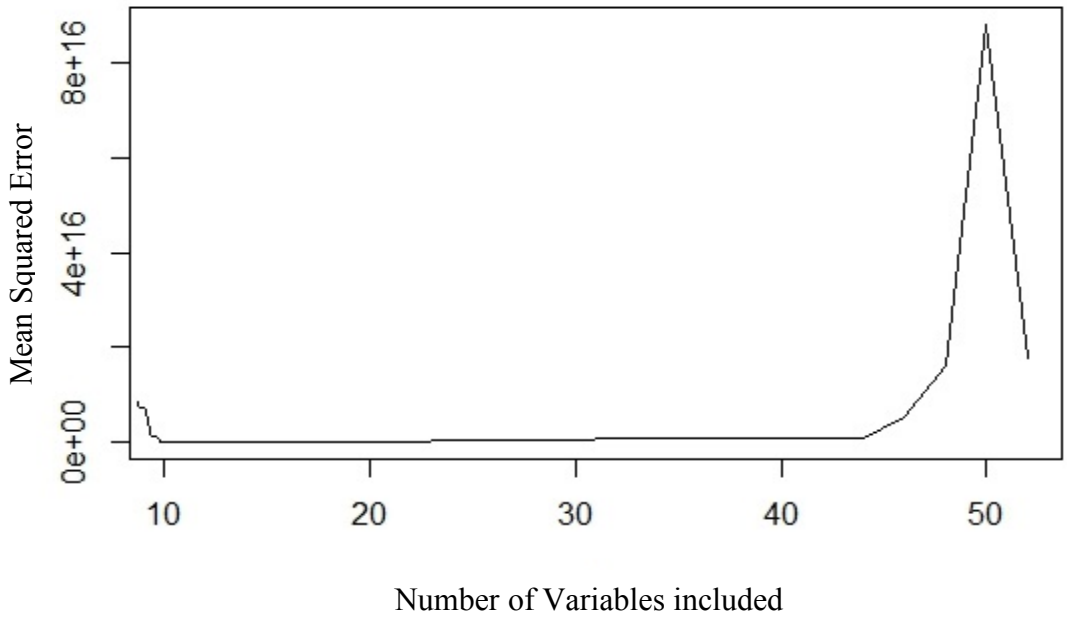

Figure 3-1 Sure independence screening

Sure Independence Screening (SIS) was used to vary the dimensionality from 0 to 70 including additive and first order interaction effects of all independent variables. Figure 3-2 shows how the mean squared error varied in relation to the number of independent variables used in Ordinary Least Squares (OLS). It was observed that the Mean Squared Error (MSE) was lowest when the 12 variables shown in Table 3-4 were used to construct the OLS model.

However, it was observed that none of the linear regression models presented statistically significant coefficient values for all the 12 variables. All linear regression methods resulted in fewer, statistically significant coefficient values. The performance of each linear regression method was compared by looking at the mean squared error as well as the adjusted $\boldsymbol{R^2}$ value.

Table 3-2 summarizes the number of statistically significant variables, MSE and adjusted $\boldsymbol{R^2}$ from all the linear models.

Table 3-2 Mean squared error from linear regression methods

Linear Regression Model	**Number of Statistically Significant variables**	**Mean Squared Error**	**Adjusted $\boldsymbol{R^2}$**
Ordinary Least Squares	3	89,207,192,840,826.40	0.98
Ridge Regression	5	10,311,488,752,154.80	0.88
Lasso	2	10,117,773,502,679.50	0.96

OLS had the highest adjusted $\boldsymbol{R^2}$. It was found that the estimated counterfeit incident index had an interaction effect on diversion of drugs. Table 3-3 lists the coefficient values, confidence intervals and p-values for these statistically significant variables obtained from OLS.

Table 3-3 Results from Ordinary least squares for diversion

Independent Variables	**Estimate**	**Std. Error**	**p - value**	**90%CI lower level**	**90%CI lower level**
(Intercept)	85,537.76	84,710.32	0.32	-85,000.00	260,000.00
Unemployment	0.00	0.00	0.35	0.00	0.01
Population: GDP	0.00	0.00	0.13	0.00	0.00
Population: CII (5)	**-0.14**	**0.08**	**0.09**	**-0.31**	**0.03**
Population: CTI (4)	-0.01	0.08	0.85	-0.17	0.14
Unemployment: CII (5)	**0.72**	**0.23**	**0.00**	**0.26**	**1.20**
Unemployment: CTI (4)	0.02	0.07	0.79	-0.12	0.16
Women: CII (5)	0.15	0.14	0.28	-0.13	0.43
Women: CTI (4)	0.02	0.15	0.91	-0.29	0.32
CII (5): CTI (3)	**25,467,942.92**	**1,372,646.85**	**0.00**	**23,000,000.00**	**28,000,000.00**

The 3 statistically significant variables at a significance level of 0.1 are all first order interaction effects who have been highlighted in Table 3-3. Interaction effects of CII (5) with population

and unemployment as well as the interaction effects of CII (5) and CTI (3) were statistically significant.

Table 3-4 shows the coefficient values obtained through lasso when the value of lambda was 401,000.

Table 3-4 Results from lasso for diversion

Independent variables	Coefficient Estimates
Population: GDP	1.3E-11
CII (5): CTI (3)	19,335,890.69

Figure 3-3 shows the residual vs fitted value plot from linear regression. This suggests a stronger tendency towards a non-linear relationship between diversion and the independent variables. The next section presents results obtained from nonlinear regression.

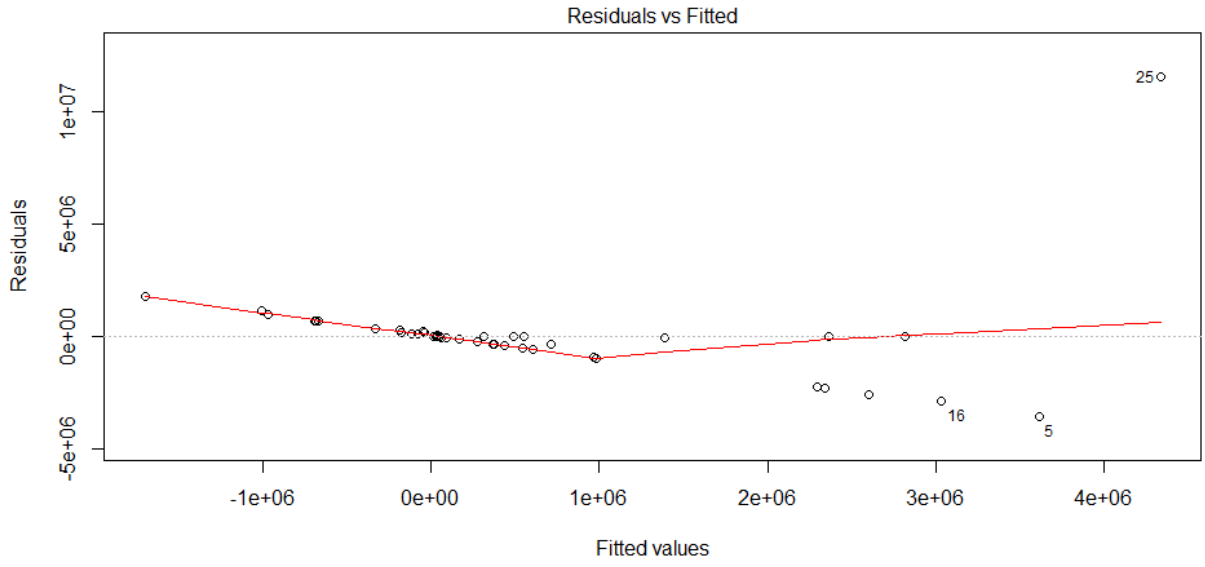

Figure 3-2 Plot of Residuals Vs Fitted values from linear regression

3.1.2 Nonlinear Regression

For the nonlinear regression, it was found that 130 trees grown with 3 variables at each split had the lowest MSE of 2,164,620,317,821.00 and explained 6.3% of variance. Figure 3-4 shows how the MSE varied with number of trees grown in the random forest

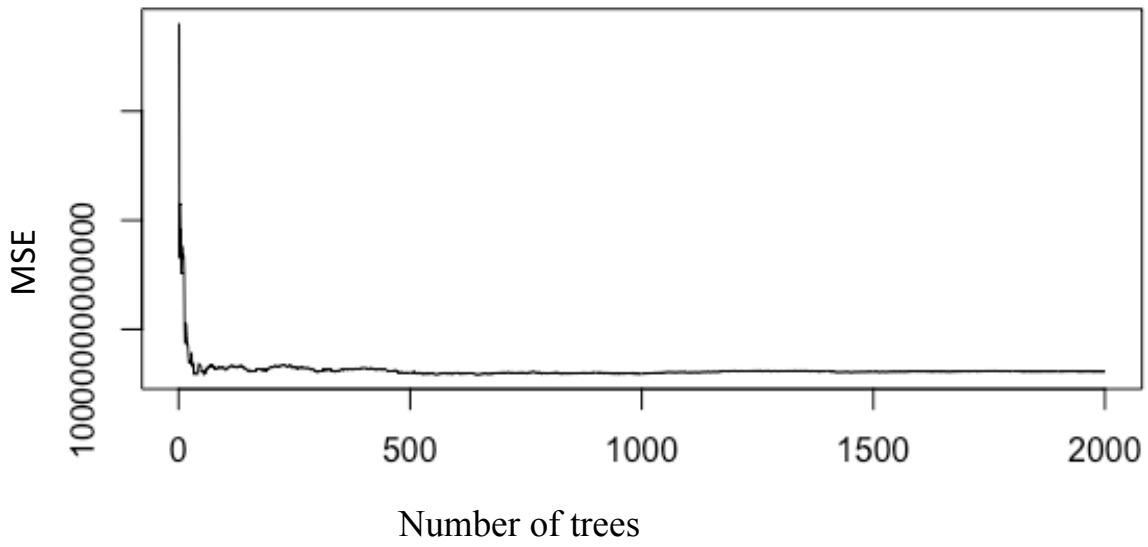

Figure 3-3 Plot of MSE vs Number of trees grown in random forest

Figure 3-5 shows the percentage increase in the MSE when the value of an independent variable is chosen from a permuted dataset instead of the training dataset. The variable whose removal results in the highest percentage increase in MSE, i.e., Death rate is the most important variable in constructing trees in the random forest.

Counterfeit incident index is the fourth least important variable as shown in Figure 3-5. This as well as results from Table 3-4 indicate that counterfeiting and diversion mainly share a linear relationship in countries which have a higher prevalence of counterfeit incidents.

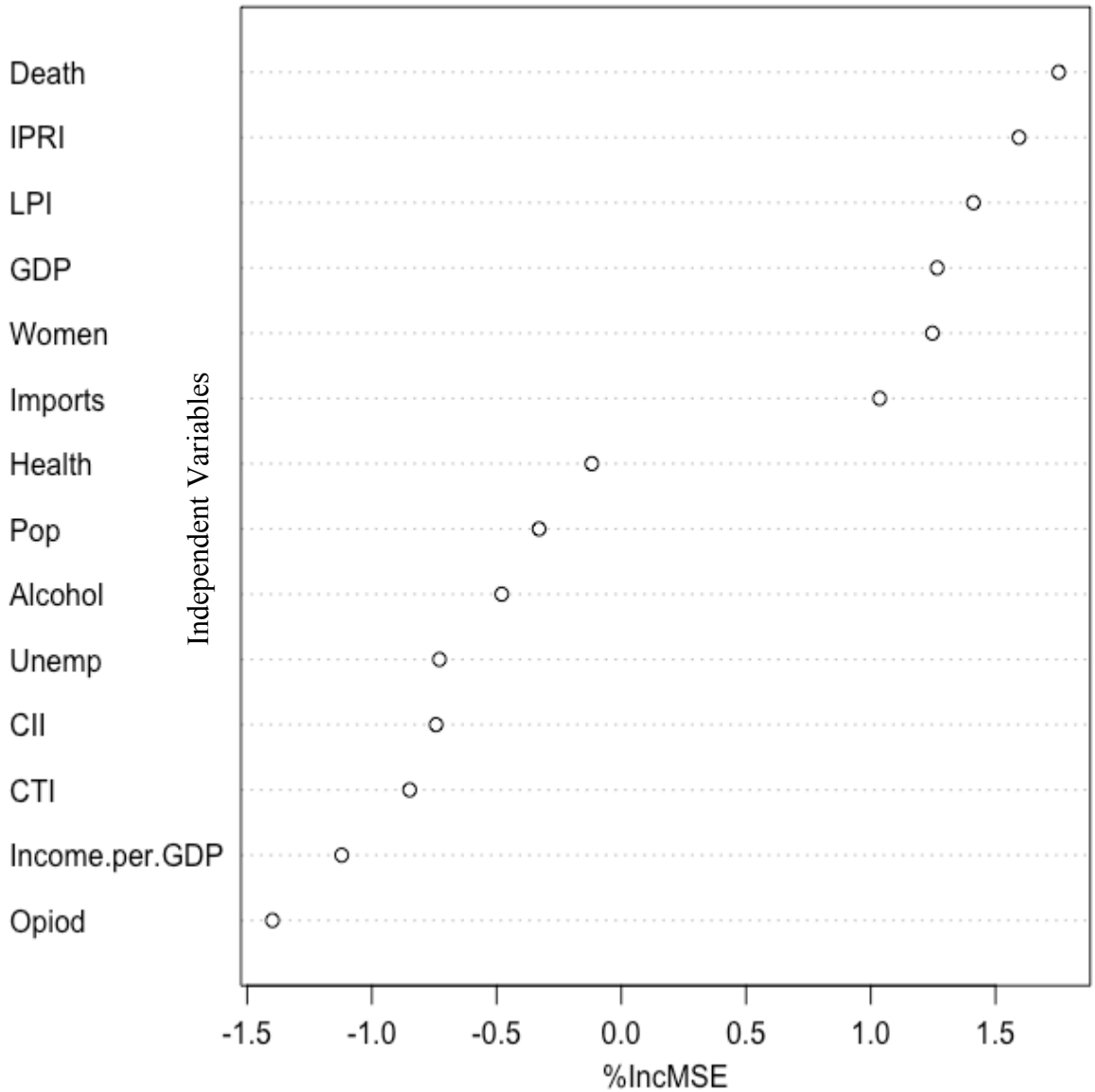

Figure 3-4 Percent increase in MSE between original and permuted dataset values for each independent variable

Figure 3-6 shows variable importance relative to other independent variables. Random forests, unlike linear regression provide only a single measure of variable importance for independent categorical variables. The variable importance scores are aggregate measures which only provide information regarding the impact of an independent variable instead of its effect on the response.

A negative value of variable importance for opioid use prevalence indicates that it has little or no impact on diversion. From Figure 3-5, the variables which have highest impact on diversion are Death, IPRI, LPI, GDP, women, quantity of pharmaceutical imports. However, the coefficient values of these most of these variables except GDP were not statistically significant in the linear model constructed. Thus, these variables have a non-linear relationship with diversion.

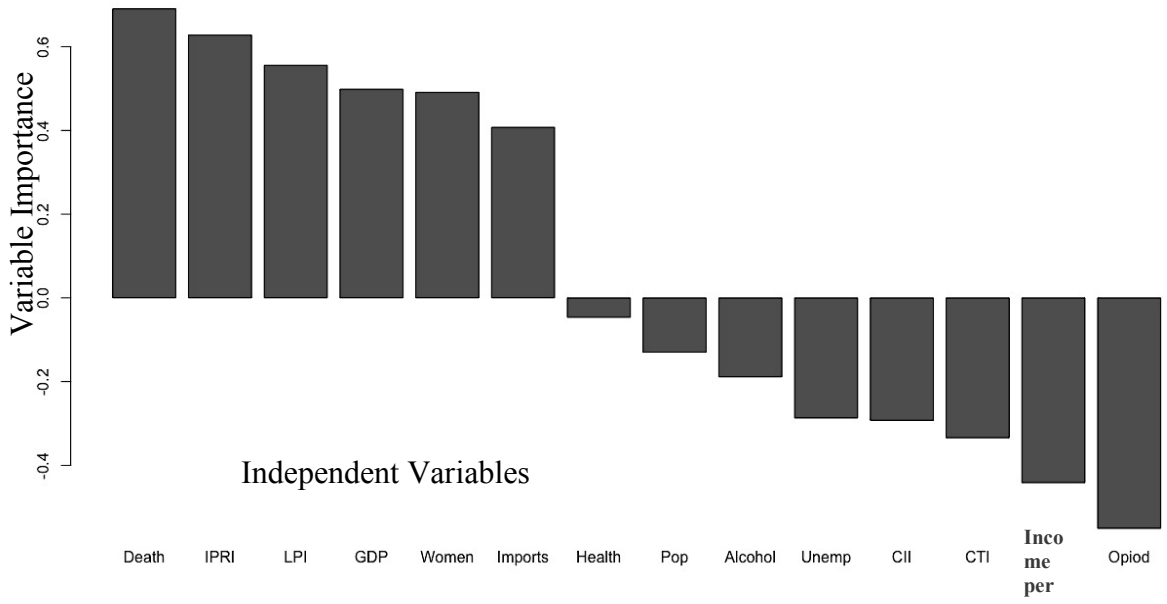

Figure 3-5 Relative variable importance in random forest model

The random forest model performed better than the linear regression methods in terms of MSE. However, this model could only explain 6.3% variance. Since MSE was the main criteria used to determine the performance of a model, the random forest model was used to predict values for diversion which are discussed further in Section 3.3.

3.2 Counterfeiting

Ordinal logistic regression resulted in the coefficient estimates shown in table 3.5

Table 3-5 Coefficient estimates from Ordinal logistic regression

Variable	Coefficient Estimate	Standard Error	p-value	90% CI (lower limit)	90% CI (upper limit)
(Intercept):1	-2.82	5.63	0.62	-13.86	8.22
(Intercept):2	-1.59	5.62	0.78	-12.61	9.42
(Intercept):3	1.23	5.66	0.83	-9.87	12.32
(Intercept):4	1.61	5.67	0.78	-9.51	12.73
(Intercept):5	2.54	5.72	0.66	-8.67	13.75
Opioid	0.10	0.64	0.88	-1.15	1.35
Population	0.00	0.00	0.04	0.00	0.00
Unemployment	0.08	0.08	0.34	-0.08	0.24
Women	0.10	0.09	0.26	-0.07	0.27
GDP	0.00	0.00	0.68	0.00	0.00
Health	0.06	0.19	0.75	-0.31	0.44
Alcohol	-0.16	0.11	0.15	-0.37	0.06
CPI	0.74	0.32	0.02	0.11	1.38
Death	-0.11	0.22	0.63	-0.54	0.33
Diversion	0.0002	0.000003	0.03	0.00	0.0004
LPI	-2.00	1.33	0.13	-4.61	0.62
CTI (1)	1.15	1.82	0.53	-2.42	4.72
CTI (2)	-2.27	1.41	0.11	-5.03	0.49
CTI (3)	2.13	1.35	0.12	-0.52	4.78
CTI (4)	-0.11	0.74	0.88	-1.56	1.34
Income per GDP	-0.01	0.01	0.41	-0.04	0.02
Imports	0.00	0.00	0.33	0.00	0.00

None of the pairwise interaction terms had statistically significant coefficient values and the model performed better without interaction terms. The prediction error rate with interaction effects was 44.44% while the error rate without interaction terms was 38.69%.

The classification error rate for each of the six levels is given below in Table 3-6:

Table 3-6 Classification error rate for ordinal logistic regression

Counterfeit Incident Index	Classification Error
0	0.06
1	0.20
2	0.22
3	0.37
4	0.70
5	0.83

The classification error rate appears to increase for higher levels of the Counterfeit Incident Index, this can be attributed to the skewed representation of levels in the training dataset, out of the data collected, only 20.9% observations had an incident index greater than 3.

According to the results, counterfeiting has a positive correlation with corruption perception index. Even though this goes against intuition, this can be attributed to the way in which data regarding counterfeiting has been collected. Countries with lower corruption and better logistics performance index have better incident reporting systems (69).

3.3 Prediction

The results from predicting estimated amount of diversion through random forest and estimated counterfeiting incident index through ordinal logistic regression are shown in Tables 3.8 and 3.9. Since the quantity of Alprazolam being transported through a country depends on the country's position on the route, the diversion estimates were predicted based on the quantity being transported for all positions of the country on the route. It was observed that the estimated amount of diversion varied based on the quantity of Alprazolam being transported while the estimated counterfeit index remained the same.

Table 3-7 Diversion estimates from predictive modeling

Position on the route	Country	Quantity of Alprazolam being transported (in USD)	Estimated diversion quantity (in Kg)
10	Bangkok, Thailand	506,200,000.00	1,802,621.79
9	Bangkok, Thailand	1,012,400,000.00	1,629,742.36
8	Bangkok, Thailand	1,518,600,000.00	1,651,037.29
7	Bangkok, Thailand	2,024,800,000.00	1,651,037.29
6	Bangkok, Thailand	2,531,000,000.00	1,651,037.29
5	Bangkok, Thailand	3,037,200,000.00	1,650,123.85
4	Bangkok, Thailand	3,543,400,000.00	1,651,430.25
3	Bangkok, Thailand	4,049,600,000.00	1,664,847.25
2	Bangkok, Thailand	4,555,800,000.00	1,647,886.94
1	Bangkok, Thailand	5,062,000,000.00	1,647,886.94
10	Da Nang, Vietnam	380,000,000.00	1,675,916.00
9	Da Nang, Vietnam	760,000,000.00	1,521,248.53
8	Da Nang, Vietnam	1,140,000,000.00	1,523,548.97
7	Da Nang, Vietnam	1,520,000,000.00	1,535,766.22
6	Da Nang, Vietnam	1,900,000,000.00	1,535,766.22
5	Da Nang, Vietnam	2,280,000,000.00	1,535,766.22
4	Da Nang, Vietnam	2,660,000,000.00	1,535,766.22
3	Da Nang, Vietnam	3,040,000,000.00	1,565,294.64
2	Da Nang, Vietnam	3,420,000,000.00	1,566,044.43
1	Da Nang, Vietnam	3,800,000,000.00	1,566,044.43
10	Guangzhou, China	450,000,000.00	1,776,214.04
9	Guangzhou, China	900,000,000.00	1,602,811.76
8	Guangzhou, China	1,350,000,000.00	1,606,344.56
7	Guangzhou, China	1,800,000,000.00	1,596,989.28
6	Guangzhou, China	2,250,000,000.00	1,596,989.28
5	Guangzhou, China	2,700,000,000.00	1,596,989.28
4	Guangzhou, China	3,150,000,000.00	1,596,075.83
3	Guangzhou, China	3,600,000,000.00	1,596,209.20
2	Guangzhou, China	4,050,000,000.00	1,611,479.96
1	Guangzhou, China	4,500,000,000.00	1,594,519.65
10	Hanoi, Vietnam	380,000,000.00	1,675,916.00
9	Hanoi, Vietnam	760,000,000.00	1,521,248.53
8	Hanoi, Vietnam	1,140,000,000.00	1,523,548.97
7	Hanoi, Vietnam	1,520,000,000.00	1,535,766.22

Table 3-8 continued

6	Hanoi, Vietnam	1,900,000,000.00	1,535,766.22
5	Hanoi, Vietnam	2,280,000,000.00	1,535,766.22
4	Hanoi, Vietnam	2,660,000,000.00	1,535,766.22
3	Hanoi, Vietnam	3,040,000,000.00	1,565,294.64
2	Hanoi, Vietnam	3,420,000,000.00	1,566,044.43
1	Hanoi, Vietnam	3,800,000,000.00	1,566,044.43
10	Ho Chi Minh City, Vietnam	380,000,000.00	1,675,916.00
9	Ho Chi Minh City, Vietnam	760,000,000.00	1,521,248.53
8	Ho Chi Minh City, Vietnam	1,140,000,000.00	1,523,548.97
7	Ho Chi Minh City, Vietnam	1,520,000,000.00	1,535,766.22
6	Ho Chi Minh City, Vietnam	1,900,000,000.00	1,535,766.22
5	Ho Chi Minh City, Vietnam	2,280,000,000.00	1,535,766.22
4	Ho Chi Minh City, Vietnam	2,660,000,000.00	1,535,766.22
3	Ho Chi Minh City, Vietnam	3,040,000,000.00	1,565,294.64
2	Ho Chi Minh City, Vietnam	3,420,000,000.00	1,566,044.43
1	Ho Chi Minh City, Vietnam	3,800,000,000.00	1,566,044.43
10	Hong Kong	506,200,000.00	1,622,274.63
9	Hong Kong	1,012,400,000.00	1,449,359.91
8	Hong Kong	1,518,600,000.00	1,441,070.92
7	Hong Kong	2,024,800,000.00	1,441,123.90
6	Hong Kong	2,531,000,000.00	1,441,123.90
5	Hong Kong	3,037,200,000.00	1,439,769.75
4	Hong Kong	3,543,400,000.00	1,439,903.12
3	Hong Kong	4,049,600,000.00	1,454,770.38
2	Hong Kong	4,555,800,000.00	1,437,756.86
1	Hong Kong	5,062,000,000.00	1,437,756.86
10	Kuala Lumpur, Malaysia	506,200,000.00	1,547,470.59
9	Kuala Lumpur, Malaysia	1,012,400,000.00	1,374,555.87
8	Kuala Lumpur, Malaysia	1,518,600,000.00	1,362,677.85
7	Kuala Lumpur, Malaysia	2,024,800,000.00	1,362,730.83
6	Kuala Lumpur, Malaysia	2,531,000,000.00	1,362,730.83
5	Kuala Lumpur, Malaysia	3,037,200,000.00	1,361,817.39
4	Kuala Lumpur, Malaysia	3,543,400,000.00	1,361,950.76
3	Kuala Lumpur, Malaysia	4,049,600,000.00	1,381,134.22
2	Kuala Lumpur, Malaysia	4,555,800,000.00	1,364,120.69
1	Kuala Lumpur, Malaysia	5,062,000,000.00	1,364,120.69
10	Kunming, China	450,000,000.00	1,776,214.04
9	Kunming, China	900,000,000.00	1,602,811.76

Table 3-8 continued

8	Kunming, China	1,350,000,000.00	1,606,344.56
7	Kunming, China	1,800,000,000.00	1,596,989.28
6	Kunming, China	2,250,000,000.00	1,596,989.28
5	Kunming, China	2,700,000,000.00	1,596,989.28
4	Kunming, China	3,150,000,000.00	1,596,075.83
3	Kunming, China	3,600,000,000.00	1,596,209.20
2	Kunming, China	4,050,000,000.00	1,611,479.96
1	Kunming, China	4,500,000,000.00	1,594,519.65
10	Phnom Penh, Cambodia	400,000,000.00	1,835,048.29
9	Phnom Penh, Cambodia	800,000,000.00	1,673,801.82
8	Phnom Penh, Cambodia	1,200,000,000.00	1,679,710.93
7	Phnom Penh, Cambodia	1,600,000,000.00	1,676,148.90
6	Phnom Penh, Cambodia	2,000,000,000.00	1,676,201.88
5	Phnom Penh, Cambodia	2,400,000,000.00	1,676,201.88
4	Phnom Penh, Cambodia	2,800,000,000.00	1,676,201.88
3	Phnom Penh, Cambodia	3,200,000,000.00	1,674,847.72
2	Phnom Penh, Cambodia	3,600,000,000.00	1,685,930.21
1	Phnom Penh, Cambodia	4,000,000,000.00	1,700,797.47
10	Vientiane, Laos	380,000,000.00	1,950,798.32
9	Vientiane, Laos	760,000,000.00	1,785,780.81
8	Vientiane, Laos	1,140,000,000.00	1,791,575.32
7	Vientiane, Laos	1,520,000,000.00	1,787,152.52
6	Vientiane, Laos	1,900,000,000.00	1,787,152.52
5	Vientiane, Laos	2,280,000,000.00	1,787,152.52
4	Vientiane, Laos	2,660,000,000.00	1,787,152.52
3	Vientiane, Laos	3,040,000,000.00	1,786,239.08
2	Vientiane, Laos	3,420,000,000.00	1,797,321.56
1	Vientiane, Laos	3,800,000,000.00	1,797,321.56

Table 3-8 Estimated Counterfeit Incident Index from predictive modeling

Country	Estimated Counterfeit Incident Index
Kuala Lumpur, Malaysia	0
Bangkok, Thailand	0
Vientiane, Laos	0
Phnom Penh, Cambodia	2
Ho Chi Minh City, Vietnam	2
Da Nang, Vietnam	2
Hanoi, Vietnam	2
Guangzhou, China	3
Kunming, China	3
Hong Kong	0

3.4 Expected costs

Expected cost of diversion, counterfeiting and transportation were calculated using the formulas mentioned in Section 2.3. Due to the huge number of possible routes (3628800), only expected costs of the existing route and the best route obtained by Dijkstra's algorithm are presented below in Tables 3-10 to 3-11.

Table 3-10 contains costs involved with the existing route used by Agility. The Source and destination represent nodes connected by a single arc. The presence of an arc sharing the same source and destination indicates a parallel arc. Each of these parallel arcs have a different mode of transportation. Parallel arcs which are more expensive were eliminated and not used in the input to Dijkstra's algorithm.

Table 3-9 Expected costs for the current route used by Agility

Source	Destination	Distance on each mode of transport			E [Cost of transportation] (in USD)	E [Cost of diversion] (in USD)	E [Cost of counterfeiting] (in USD)
		Road (km)	Water (km)	Air (km)			
Singapore	Kuala Lumpur	354.76			865.61	2.07E+12	0.00
Singapore	Kuala Lumpur		446.06		749.38	2.76E+11	0.00
Singapore	Kuala Lumpur			354.76	4,328.07	1.38E+11	0.00
Kuala Lumpur	Bangkok			1237	15,091.40	1.67E+11	0.00
Kuala Lumpur	Bangkok	1471.93			3,591.51	2.50E+12	0.00
Kuala Lumpur	Bangkok		1952.13		3,279.58	3.34E+11	0.00
Bangkok	Vientiane			536.59	6,546.40	1.36E+11	0.00
Bangkok	Vientiane	646.47			1,577.39	2.04E+12	0.00
Bangkok	Phnom Penh			563.53	6,875.07	1.34E+11	7,853,394,028.00
Bangkok	Phnom Penh	659.38			1,608.89	2.01E+12	9,650,882,807.00
Bangkok	Phnom Penh		1150.54		1,932.91	2.68E+11	956,448,504.20
Phnom Penh	Ho Chi Minh City			230.1	2,807.22	1.17E+11	7,188,793,885.00
Phnom Penh	Ho Chi Minh City	279.79			682.69	1.75E+12	4,131,020,166.00
Phnom Penh	Ho Chi Minh City		941.87		1,582.34	2.33E+11	4,439,228,582.00
Ho Chi Minh City	Da Nang			615.71	7,511.66	1.17E+11	6,033,745.83
Ho Chi Minh City	Da Nang	858.64			2,095.08	1.75E+12	10,001,006.82
Ho Chi Minh City	Da Nang		2374.56		3,989.26	2.33E+11	1,531,791,635.00
Da Nang	Hanoi			659.91	8,050.90	1.16E+11	4,530,504,827.00
Da Nang	Hanoi	767.76			1,873.33	1.74E+12	9,685,185,552.00
Da Nang	Hanoi		956.88		1,607.56	2.32E+11	5,861,701,220.00
Hanoi	Guangzhou			831.01	10,138.32	1.44E+11	13,169,839,005.00
Hanoi	Guangzhou	801.24			1,955.03	2.16E+12	37,143,152,787.00
Hanoi	Guangzhou		1084.85		1,822.55	2.89E+11	35,182,146,689.00
Vientiane	Kunming			818.27	9,982.89	1.60E+11	36,562,938,755.00
Vientiane	Kunming	850			2,074.00	2.40E+12	40,211,332,789.00
Guangzhou	Hong Kong			147.17	1,795.47	1.64E+11	0.00
Guangzhou	Hong Kong	119			290.36	2.46E+12	0.00
Guangzhou	Hong Kong		182.52		306.63	3.28E+11	0.00

Expected total cost was obtained by using equation 2.11 to form the cost matrix. Table 3-12 shows a part of the cost matrix used as input for Dijkstra's algorithm. The cells remaining blank indicate that there is no arc connecting the two countries directly. This matrix does not contain any parallel arcs and represents a simple graph. To make the cost matrix more readable, the matrix labels have been renamed to represent the route. Table 3-11 is the key to the cost matrix.

Table 3-10 Key to cost matrix

Country	Matrix labels for route 1	Matrix labels for route 2
Singapore	A1	A2
Kuala Lumpur, Malaysia	B1	B2
Bangkok, Thailand	C1	C2
Vientiane, Laos	D1	D2
Phnom Penh, Cambodia	E1	E2
Ho Chi Minh City, Vietnam	F1	F2
Da Nang, Vietnam	G1	G2
Hanoi, Vietnam	H1	H2
Guangzhou, China	I1	I2
Kunming, China	J1	J2
Hong Kong	K1	K2

Table 3-11 Cost matrix for routes 1 and 2

	A1	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	A2	B2	C2	D2	E2	F2	G2	H2	I2	J2	K2
A1									5E+10													
B1			4E+10																			
C1				3E+10																		
D1					7E+10																	
E1						3E+10																
F1							3E+10															
G1								4E+10														
H1																						
I1										4E+10												
J1											4E+10											
K1		4E+10											4E+10									
A2																				5E+10		
B2															4E+10							
C2																						
D2																	3E+10					
E2																		3E+10				
F2																			3E+10			
G2														4E+10								
H2																					4E+10	
I2																						4E+10
J2																3E+10						

3.5 Best route

Expected cost of the best route obtained was USD 373381606670.791. Figure 3.7 presents the best route obtained in red, all other possible routes are shown in grey color. The results are tabulated below in Table 3-13.

The total expected cost of the current route is USD 20971455104097.5. This cost is higher than the expected cost of the best route chosen by Dijkstra's algorithm by USD 20598073497426.7. Mode of transportation chosen for the best route was air. Despite being the most expensive mode of transportation, the probability of diversion occurring is the least with this mode. This is due to the higher weight assigned to the occurrence of diversion compared to the cost of transportation. The expected cost of transportation for the existing route with the mode of transportation being road is USD 53483.78. This is much lower than the expected cost of transportation for the route chosen by Dijkstra's algorithm which is USD 124180.5, about 2.32 times the cost of transportation for the existing route.

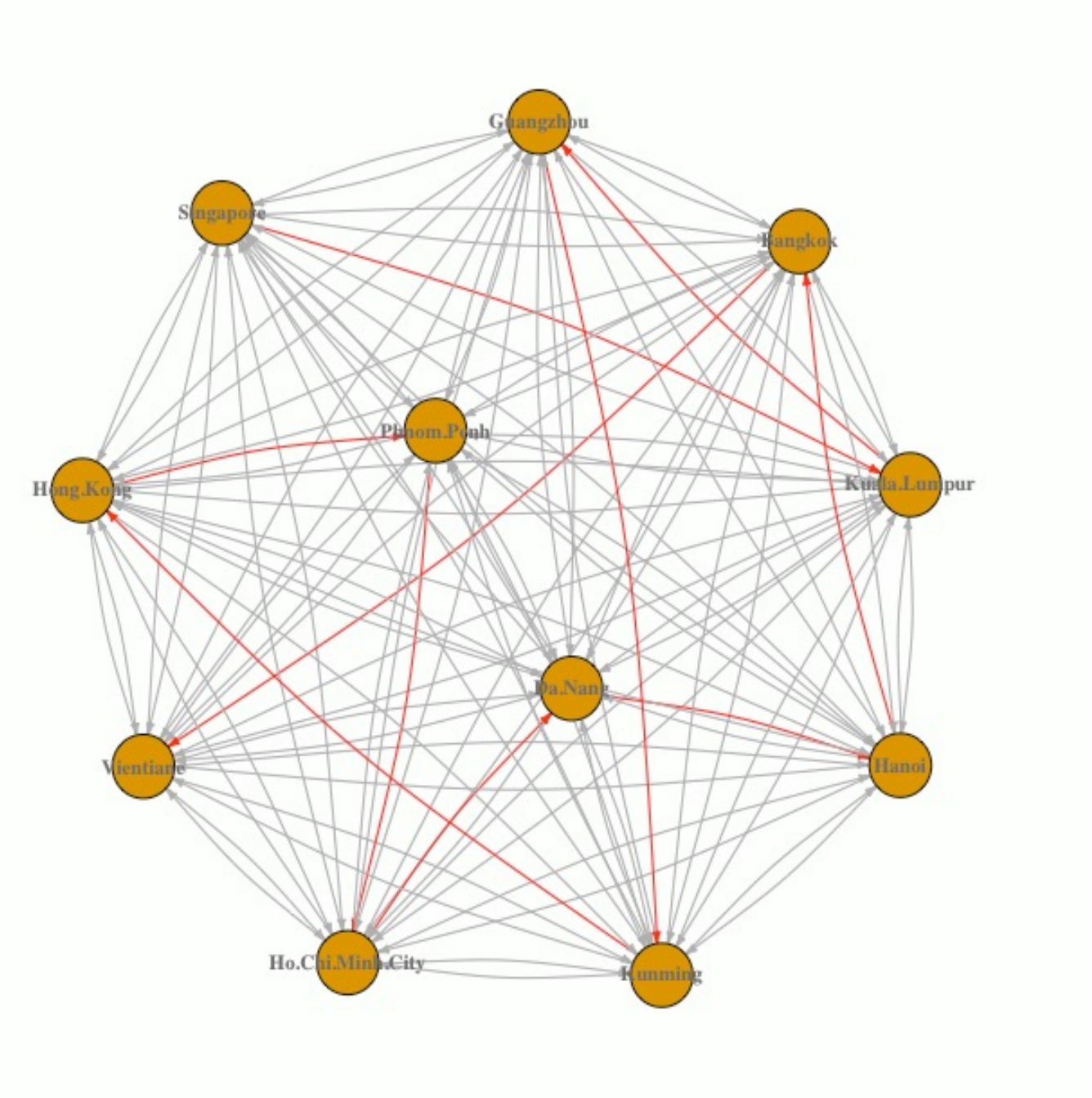

Figure 3-6 Best route obtained through Dijkstra's algorithm

Table 3-12 Costs associated with the best route chosen by Dijkstra's algorithm

Source	Destination	Distance	Mode of transport	E [Cost of transportation]	Diversion estimate	E [Cost of diversion]	Estimated Counterfeit Incident Index	E [Cost of Counterfeiting]	E [Total Arc Cost]
Singapore	Kuala Lumpur	327.01	Air	3,989.52	1,364,120.69	34,525,894,664.00	0.00	0.00	34,525,898,653.00
Kuala Lumpur	Guangzhou	2,558.92	Air	31,218.82	1,611,479.96	36,258,299,100.00	3.00	11,608,571,346.00	47,866,901,665.00
Guangzhou	Kunming	1,121.91	Air	13,687.30	1,596,075.83	35,911,706,175.00	3.00	5,948,359,187.00	41,860,079,049.00
Kunming	Hong Kong	1,516.97	Air	18,507.03	1,439,903.12	36,443,947,967.00	0.00	0.00	36,443,966,474.00
Hong Kong	Phnom Penh	1,574.19	Air	19,205.12	1,676,201.88	33,524,037,600.00	2.00	2,600,705,151.00	36,124,761,956.00
Phnom Penh	Ho Chi Minh City	230.10	Air	2,807.22	1,535,766.22	29,179,558,180.00	2.00	3,095,803,863.00	32,275,364,850.00
Ho Chi Minh City	Da Nang	615.71	Air	7,511.66	1,535,766.22	29,179,558,180.00	2.00	4,226,232,986.00	33,405,798,678.00
Da Nang	Hanoi	659.91	Air	8,050.90	1,523,548.97	28,947,430,430.00	2.00	3,617,430,450.00	32,564,868,931.00
Hanoi	Bangkok	1,037.42	Air	12,656.52	1,629,742.36	41,248,779,132.00	0.00	0.00	41,248,791,788.00
Bangkok	Vientiane	536.59	Air	6,546.40	1,950,798.32	37,065,168,080.00	0.00	0.00	37,065,174,626.00

This route however was found to be the best route even when the preferred mode of transportation was road. Table 3-13 shows the cost associated with this route when road is preferred over air. The total expected cost for this route was USD 3,592,926,654,459.82, which is higher than the expected cost for the best route by USD 20,598,073,497,426.7 but still lower than the expected cost for the existing route. The expected cost of transportation for this route was USD 76,431.63 which is 1.42 times the current expected transportation cost.

The next section shows how these results varied when sensitivity analysis was used to test the robustness of the models developed.

Table 3-13 Expected arc costs when preferred mode of transportation is road

Source	Destination	Distance	Mode of transport	E [Cost of transportation]	Diversion estimate	E [Cost of diversion]	Estimated Counterfeit Incident Index	E [Cost of Counterfeiting]	E[Total Arc Cost]
Singapore	Kuala Lumpur	354.76	Road	865.61	517,888,000,000.00	34,525,894,664.00	0.00	0.00	521,544,000,000.00
Kuala Lumpur	Guangzhou	2,558.92	Air	31,218.82	36,258,299,100.00	36,258,299,100.00	3.00	11,608,571,346.00	38,891,591,861.00
Guangzhou	Kunming	1,372.37	Road	3,348.58	538,676,000,000.00	35,911,706,175.00	3.00	5,948,359,187.00	538,676,000,000.00
Kunming	Hong Kong	1,192.00	Road	2,908.48	546,659,000,000.00	36,443,947,967.00	0.00	0.00	546,659,000,000.00
Hong Kong	Phnom Penh	1,574.19	Air	19,205.12	33,524,037,600.00	33,524,037,600.00	2.00	2,600,705,151.00	33,524,056,805.00
Phnom Penh	Ho Chi Minh City	279.79	Road	682.69	437,693,000,000.00	29,179,558,180.00	2.00	3,095,803,863.00	437,693,000,000.00
Ho Chi Minh City	Da Nang	858.64	Road	2,095.08	437,693,000,000.00	29,179,558,180.00	2.00	4,226,232,986.00	441,129,000,000.00
Da Nang	Hanoi	767.76	Road	1,873.33	434,211,000,000.00	28,947,430,430.00	2.00	3,617,430,450.00	435,504,000,000.00
Hanoi	Bangkok	659.91	Air	12,656.52	41,248,779,132.00	41,248,779,132.00	0.00	0.00	43,065,056,506.00
Bangkok	Vientiane	646.47	Road	1,577.39	555,978,000,000.00	37,065,168,080.00	0.00	0.00	556,241,000,000.00

4 Sensitivity Analysis

Sensitivity analysis was used to determine the robustness of the model and to understand how uncertainties in the data used for modeling can affect the results obtained. Table 4.1 presents the variable of interest for sensitivity analysis and the source of uncertainty.

Table 4-1 Variable used for conducting sensitivity analysis

Variable	Base Case range	Source of uncertainty
Diversion estimate from prescription drug seizures	0 – 23,000,000	Percentage of diversion only from the distribution chain is not specified

Prescription drug seizures were used as an estimate for amount of diversion in the regression models. However, the actual diversion from the pharmaceutical distribution chain might be higher or lower. This was tested by varying the value of diversion in steps of 1% the prescription drug seizure estimate. These values were first used in the regression model to test the effect of the sensitivity analysis on the value of the coefficients.

The new diversion estimate was used as the response in random forest to test for sensitivity of the important variables. It was found that important variables remained the same when diversion estimate was between 95% to 115% of the prescription drug seizure estimates. Figures 4-1 and 4-2 display the relative variable importance at 94% and 116% of the prescription drugs seizure estimate. There were more number of variables having a positive importance value in both the cases. Unemployment, which was not considered to have importance in the model first constructed shows a positive importance value in both Figures 4-1 and 4-2.

The change in variable importance also affected the shortest paths chosen from Dijkstra's algorithm. Figures 4-3 and 4-4 show this change in shortest paths. This change in best route

indicates how the weight assigned to a factor which might affect diversion can lead to a different outcome.

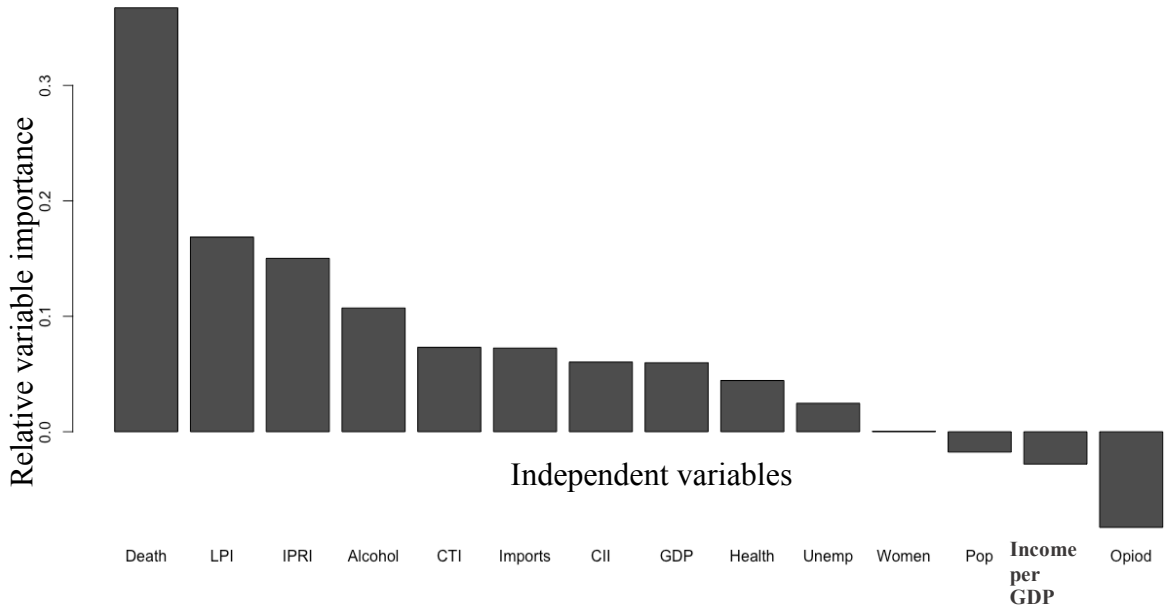

Figure 4-1 Relative variable importance when diversion = 94% prescription drug seizure estimate

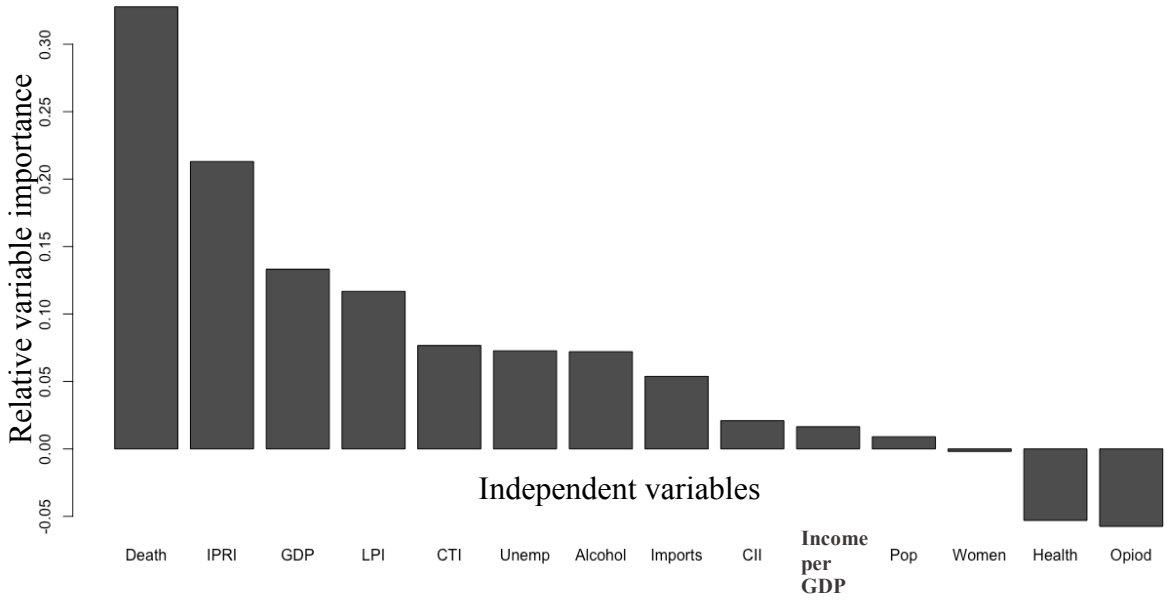

Figure 4-2 Relative variable importance when diversion=116% prescription drug seizure estimate

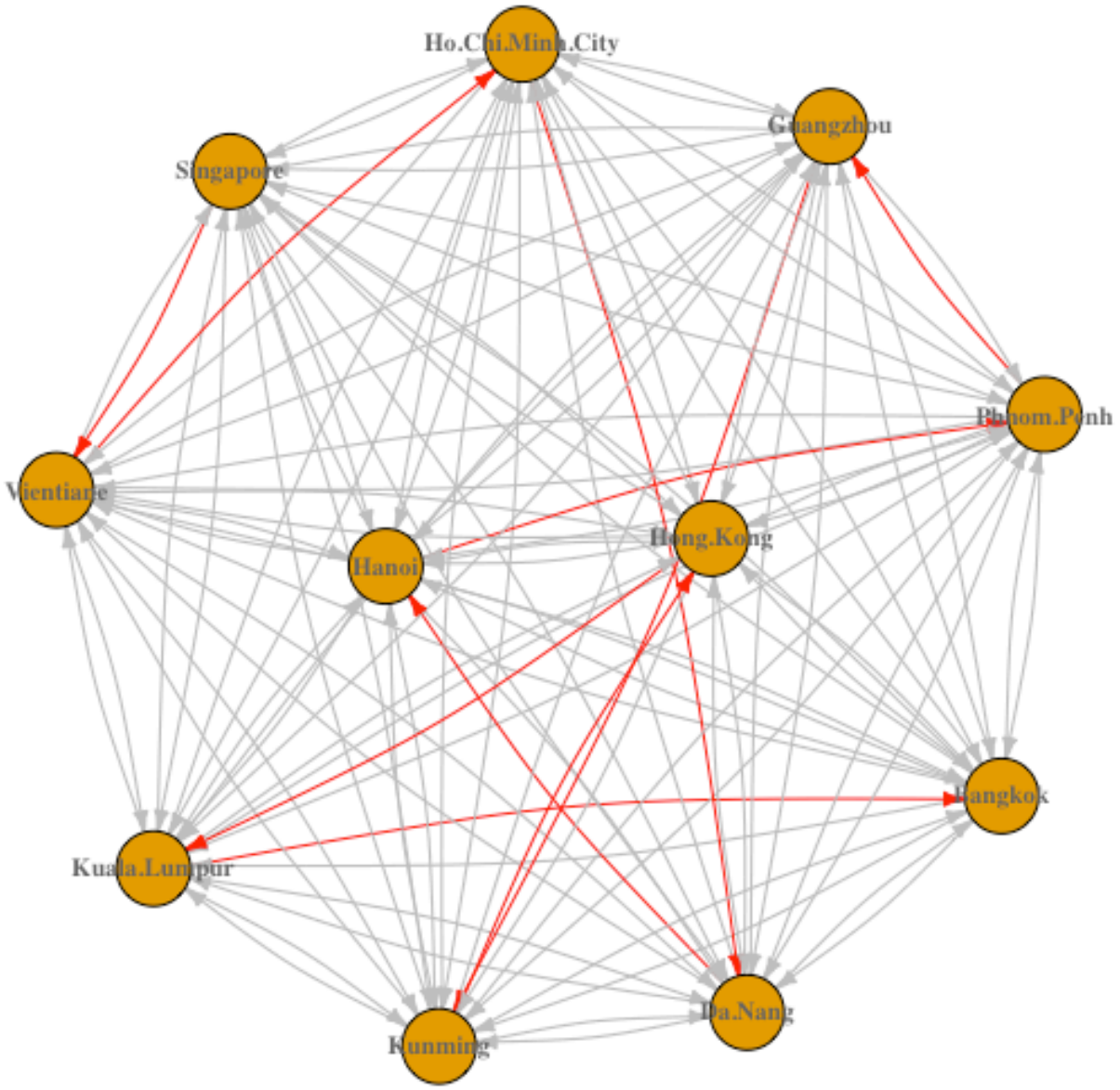

Figure 4-3 Best route when diversion=94% prescription drug seizure estimate

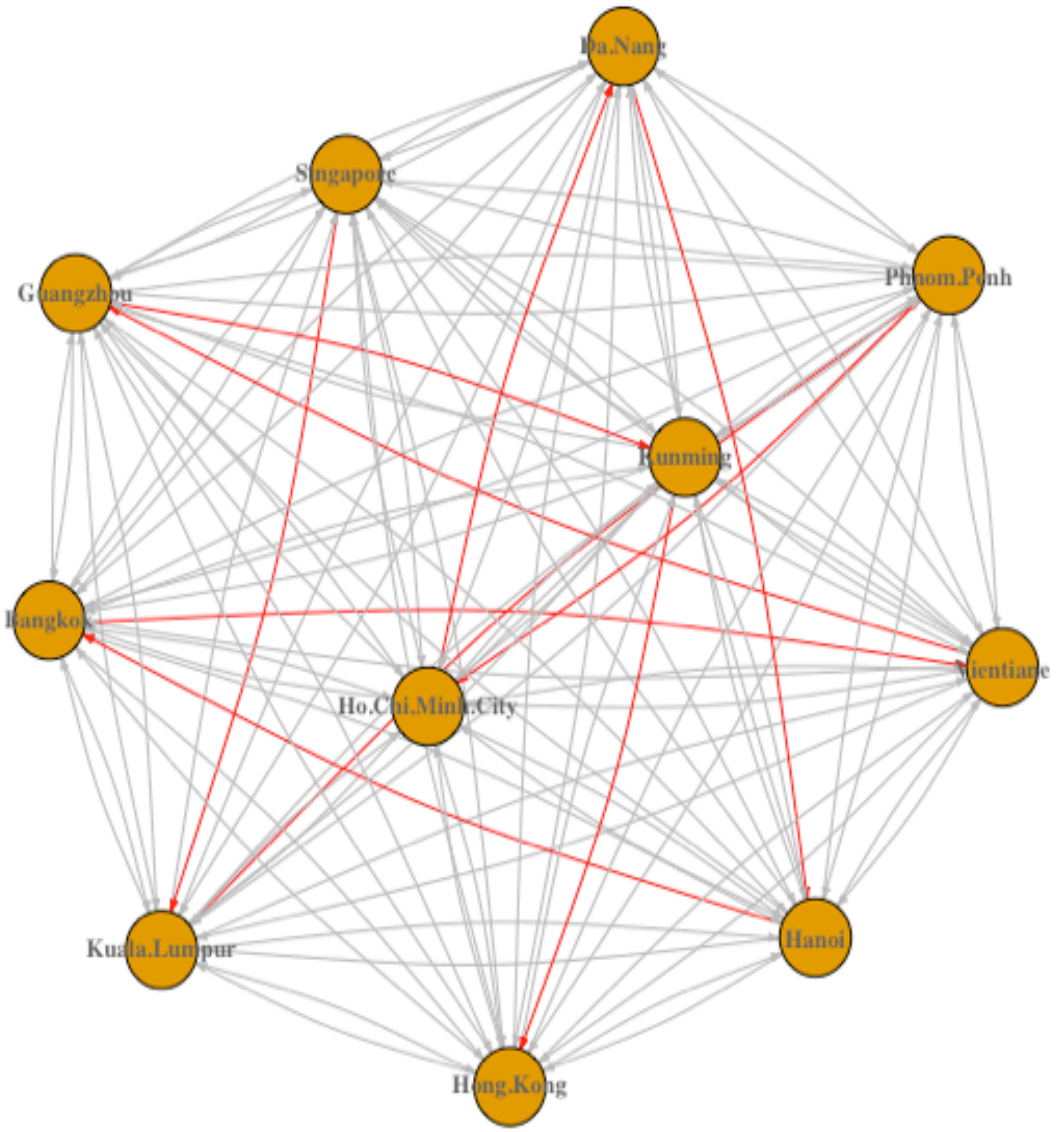

Figure 4-4 Best route when diversion=116% prescription drug seizure estimate

These changes in best route obtained indicate how sensitive the model is to the changes in data used to understand the relationship between diversion and the independent variables.

5 Conclusion

Drug diversion and counterfeiting pose a serious threat to the quality of the pharmaceutical supply chain. The outcome of these crimes negatively affect the economy as well as human lives. The objectives of this thesis were to

1. Study factors which have previously been considered to affect diversion and counterfeiting and to quantify the relationship between the factors, diversion and counterfeiting.
2. Develop a model which can predict the cost of diversion and counterfeiting and use the predicted values to find the best route to transport pharmaceutical products globally

Information regarding 16 variables of interest, diversion and counterfeiting were used to develop linear and nonlinear regression models with diversion as the dependent variable and an ordinal logistic regression model with counterfeiting as the dependent variable. These models revealed that counterfeiting had a strong positive interaction effect with population, unemployment, and cargo theft on diversion. The ordinal logistic regression model showed that population, corruption perception index and diversion have a positive additive effect on counterfeiting.

Despite the smaller number of statistically significant variables chosen by the regression and classification models, the percentage of variance explained was high, which shows how important the variables chosen were. The high coefficient estimates of 19,335,890.69 by lasso (adjusted R^2= 0.96) and 25,467,942.92 by ordinary least squares (adjusted R^2= 0.98) for the interaction effects of Cargo theft index with counterfeit incident index on diversion impresses upon the importance of routing to reduce the risk of diversion and counterfeiting.

It was also revealed that the logistics performance index shared a positive nonlinear relationship with diversion. Even though this goes against intuition, this relationship may be attributed to how the data regarding diversion and counterfeiting is reported. A country with a higher logistics performance index or corruption perception index may have better incidence

reporting systems in place. This leads to a data collection which is biased and might not be the true representative of diversion or counterfeiting.

Using Dijkstra's algorithm to obtain the best route showed how choosing the route or mode of transport which has the least cost of transportation might turn out to be more expensive when the cost of risks such as diversion and counterfeiting are considered. Saving cost on transportation might be beneficial in the short run, however it gets expensive in the long run due to risks involved. This model showed how costs of diversion and counterfeiting can be reduced even with a preferred mode of transportation which might be risky but is cheaper if the factors affecting diversion and counterfeiting are considered while planning the route. This further indicates the importance of using route analysis along with techniques such as RFID and barcoding which are suggested in the literature.

Sensitivity analysis showed how sensitive the model is to changes since the consequences of diversion and counterfeiting are considered to be higher than the cost of transportation. Relationship between independent and dependent variables also depends on the quality and quantity of data used in modelling.

The major limitation encountered during this thesis was the lack of data regarding diversion and counterfeiting. This limitation has also been brought to light in previous literature and has been one of the main reasons for the lack of quantitative analysis of the factors affecting diversion and counterfeiting in literature. Improving supply chain visibility using these techniques can improve the quality of data collected regarding diversion and counterfeiting incidents.

As a continuation and improvement on this thesis, the model can be expanded to include the cost of diversion and counterfeiting in local regions within a country and use this information to provide the probability of relative regional risks. Last mile thefts have been increasing in the recent years. A model which can provide relative regional risks can help plan the safe delivery of pharmaceutical goods in the last mile.

Collecting better data is necessary for expanding this model to include smaller regions instead of aggregating multiple regions with varied risks of diversion and counterfeiting into a country. This will be useful in efficiently routing the transportation of pharmaceutical products and will also help improve the quality of supply chain and understand factors which cause crimes like diversion and counterfeiting to occur at a regional level.

References

1. Administration, U.S. Food and Drug. U.S. Food and Drug Administration. *U.S. Food and Drug Administration.* [Online] [Cited: 06 29, 2016.] http://www.fda.gov/Drugs/DrugSafety/DrugIntegrityandSupplyChainSecurity/.

2. PSI. Pharmaceutical Security Institute - Incident Trends. *Pharmaceutical Security Institute.* [Online] [Cited: 06 29, 2016.] psi-inc.org/incidentTrends.cfm.

3. Pharmaceutical Security Institute. Pharmaceutical Security Institute. *Pharmaceutical Security Institute.* [Online] [Cited: 06 29, 2016.] psi-inc.org/index.cfm.

4. Rx-360. Rx-360. *Rx-360.* [Online] [Cited: 06 29, 2016.] rx-360.org/en-us/About-Rx-360.

5. Chuck Forsaith, Dan Burges. *RISKS IN THE PHARMACEUTICAL SUPPLY CHAIN.* s.l. : Cargo Security International, 2009.

6. Depression and Bipolar Support Alliance. Generic andn Brand Name Drugs: Understanding the basics. [Online] [Cited: 06 29, 2016.] www.dbsalliance.org/pdfs/GenericRx.pdf.

7. FDA. What are Generic Drugs? *U.S. Food and Drug Administration.* [Online] [Cited: 05 13, 2016.] www.fda.gov/Drugs/ResourcesForYou/Consumers.

8. The Health Strategies Consultancy LLC. Understanding the U.S. Commercial Pharmaceutical Supply Chain. [Online] March 2005. [Cited: 05 27, 2016.] http://avalere.com/research/docs/Follow_the_Pill.pdf.

9. Gostin, Gillian J. Buckley and Lawrence O. *Countering the Problem of Falsified and Substandard Drugs.* s.l. : National Academy of Sciences, 2013.

10. Healthcare Distribution Alliance. *Traceability: Transforming the Pharmaceutical Supply Chain.* s.l. : Healthcare Distribution Alliance, 2016.

11. Modern Distribution Management. 2015 MDM market leaders | Top Pharmaceutical Distributors. *www.mdm.com.* [Online] [Cited: 07 23, 2016.] www.mdm.com/2015-top-pharmaceuticals-distributors.

12. U.S. Food and Drug Administration. *Profile of the prescription drug wholesaling industry.* s.l. : U.S. Food and Drug Administration.

13. *Integrity of the pharmaceutical supply chain : Product sourcing for patient safety.* HEALTH INDUSTRY GROUP PURCHASING ASSOCIATION. 18, s.l. : American Journal of Health-System Pharmacy, 2004, Vol. 61.

14. Contractpharma. Glossary. *Contractpharma.* [Online] Feb 27, 2012. [Cited: 05 8, 2016.] http://www.contractpharma.com/contents/view_glossary/2012-02-27/contract-manufacturing-organization-cmo-.

15. *Evaluating outsourcing partners' capability: a case study from the pharmaceutical supply chain.* Mehta, Min Zhang Kulwant S. Pawar Janat Shah Peeyush. 8, s.l. : Emerald Insight, 2013, Journal of Manufacturing Technology Management, Vol. 24, pp. 1080-1101.

16. *Implications of outsourcing on operations planning: findings from the pharmaceutical industry.* Youssef Boulaksil Jan C. Fransoo. 10, s.l. : Emerald Insight, 2010, International Journal of Operations & Production Management, Vol. 30, pp. 1059-1079.

17. Douglas, Merrill. Navigating Pharma Logistics. *Inbound Logistics.* [Online] Inbound Logistics, August 2012. [Cited: 06 10, 2016.] http://www.inboundlogistics.com/cms/article/navigating-pharma-logistics/.

18. Global Market Insights. Globenewswire. *Nasdaq.* [Online] Global Market Insights, May 05, 2016. [Cited: 10 19, 2016.] https://globenewswire.com/news-release/2016/05/05/836881/0/en/Third-Party-Logistics-3PL-Market-size-to-reach-1-029-47-Billion-by-2022-Global-Market-Insights-Inc.html.

19. http://www.drugwarfacts.org. [Online] http://www.drugwarfacts.org/cms/Diversion#sthash.uzJioGk0.dpuf.

20. Drug Enforcement Administration. *The Trafficking and Abuse of Prescription Controlled Substances, Legend Drugs and Over the Counter products.* 2013.

21. National Estimates of Drug-related Emergency Department Visits, 2004 - 2011 - Nonmedical use of Pharmaceuticals. *Substance Abuse and Mental Health Services Administration.* [Online] 2011. http://www.samhsa.gov/data/emergency-department-data-dawn/reports.

22. UNODC. Information about Drugs. *UNODC.* [Online] [Cited: 04 06, 2016.] https://www.unodc.org/unodc/en/illicit-drugs/definitions/.

23. Elrod, Brad. Pharmaceutical Based Cargo Security and Theft Prevention. *Pharmaceutical Technology Europe.* 2012.

24. *Collaboration or Coercion? Partnering to Divert Prescription Opioid Medications.* Traci C. Green, corresponding author Sarah E. Bowman, Madeline Ray, Nickolas Zaller, Robert Heimer, and Patricia Case. s.l. : Journal of Urban Health : Bulletin of the New York Academy of Medicine, 2013. 90(4):758-767.

25. Rx-360 Supply Chain Security Illegal Diversion. Illegal Diversion of Pharmaceuticals: How should we tackle this increasing threat to patient safety? *http://www.rx-360.org.*

[Online] April 30, 2013. http://www.rx-360.org/LinkClick.aspx?fileticket=UgkgMFRt0ak%3D&tabid=209.

26. Canadian Citizen Sentenced in Scheme to Defraud Consumers Purchasing Pharmaceuticals Online. *The United States Department of Justice.* [Online] 2013. http://www.justice.gov/opa/pr/canadian-citizen-sentenced-scheme-defraud-consumers-purchasing-pharmaceuticals-online.

27. *Drug Shortages in the US – Causes, Impact, and Strategies.* Anna Gu, Albert I. Wertheimer, Bernard Brown, Fadia T. Shaya. 4, s.l. : Innovations in Pharmacy, 2011, Vol. 2.

28. *Product safety and security in the global supply chain: Issues, challenges and research opportunities.* Ann Maruchecka, Noel Greisb, Carlos Menac, Linning Caid. 7-8, s.l. : Journal of Operations Management, 2011, Vol. 29.

29. Paeffgen, Jonathan E. Wells and Elise N. *EPA May Threaten Reverse Distribution Of Pharmaceuticals.* s.l. : LAW 360, 2014.

30. FreightWatch International. *2013 Global Cargo Theft Threat Assessment.* s.l. : FreighWatch International, 2013.

31. Three Individuals Indicted For Prescription Drug Diversion Conspiracy. *United States Department of Justice.* [Online] United States Department of Justice, January 24, 2013. http://www.justice.gov/usao-mdtn/pr/three-individuals-indicted-prescription-drug-diversion-conspiracy.

32. Palmer, Eric. Italian mafia at center of network that stole vials of Roche's Herceptin. *Fierce Pharma Manufacturing.* [Online] Fierce Pharma Manufacturing, 2014. http://www.fiercepharmamanufacturing.com/story/italian-mafia-center-network-stole-vials-roches-herceptin/2014-05-06.

33. Armstrong, David. Pharmacy delivery vans targeted by thieves seeking painkillers. *STAT.* [Online] December 22, 2015. http://www.statnews.com/2015/12/22/pharmacy-delivery-vans-targeted/.

34. Exel: Top 10 Pharma Cargo Thefts. *FiercePharma Manufacturing.* [Online] FiercePharma Manufacturing, June 29, 2010. http://www.fiercepharmamanufacturing.com/special-reports/top-10-pharma-cargo-thefts-value-2009-2010/exel-top-10-pharma-cargo-thefts.

35. Jim Jorgenson, RPh, MS, FASHP. *Controlled Substance Diversion: Tales from the front.* Indianapolis, IN : Clarian Health Partners/Indiana University Health.

36. MAINE BOARD OF PHARMACY. *Case against John J. Bartash, Jr.* 2008-PHA-4912 & PR 2805, 2008-PHA-4912 & PR 2805. Maine : Maine Supreme Court, 2011.

37. *Diversion of Drugs Within Health Care Facilities, a Multiple-Victim Crime: Patterns of Diversion, Scope, Consequences, Detection, and Prevention.* Berge, Keith H, MDView Profile, et al. s.l. : Mayo Clinic Proceedings, 2012. ISSN 00256196.

38. Jessie Balmert, Gannett Ohio. Who watches prescription drug collection? *Coshoctone Tribune.* [Online] Coshoctone Tribune, 2016. http://www.coshoctontribune.com/story/news/crime/high-in-ohio/2016/02/07/watches-prescription-drug-collection/79963112/.

39. Roy S. Fenof, Jeremy M Wilson. *Africa's Counterfeit Pharmaceutical Epidemic: The road ahead.* s.l. : A-CAPPP (Anti-Counterfeiting and Product Protection Program), 2009.

40. Lomb, Dr. Rüdiger. Securing the Global Pharmaceutical Supply Chain against the Threat of Counterfeit Drugs. *World Courier.* [Online] World Courier, June 1, 2014. http://www.worldcourier.com/insights-events/market-insights/detail/securing-global-pharmaceutical-supply-chain.

41. Ford, Neil. India, China challenge big pharmaceutical companies in Africa. *African Business Magazine.* [Online] African Business Magazine, December 10, 2014. http://africanbusinessmagazine.com/uncategorised/india-china-challenge-big-pharmaceutical-africa/.

42. C, Kapp. *Counterfeit drug problem "underestimated", says conference.* 2002.

43. *Antimalarial drug quality in Africa.* Amin A, Kokwaro GO. 5, s.l. : J Clin Pharm Ther, 2007, Vol. 32.

44. *The Health and Economic Effects of Counterfeit Drugs.* Erwin A. Blackstone, PhD, Joseph P. Fuhr, Jr, PhD and Steve Pociask, MA. 4, s.l. : American Health & Drug Benefits, 2014, Vol. 7.

45. *COUNTERFEIT DRUGS: FIGHTING ILLEGAL SUPPLY CHAINS.* U.S.HOUSE OF REPRESENTATIVES, COMMITTEE ON ENERGY AND COMMERCE, SUBCOMMITTEE ON OVERSIGHT AND INVESTIGATIONS. s.l. : U.S.HOUSE OF REPRESENTATIVES, COMMITTEE ON ENERGY AND COMMERCE, SUBCOMMITTEE ON OVERSIGHT AND INVESTIGATIONS, 2014.

46. Lew Kontnik, Director, Brand Protection, Amgen. Fake Avastin® Pulls the Curtain Back on Internet Pharmacies. *Rx-360.* [Online] Rx-360, April 2012. [Cited: 02 29, 2016.] http://rx-360.org/Portals/1/Alerts/Rx-360%20Avastin%202012-03%20r8.pdf.

47. *The Fake and the Fatal: The Consequences of Counterfeits.* LEWIS, KEVIN. s.l. : The Park Place Economist, Vol. XVII.

48. Transport Asset Protection Association. IIS_Reports. [Online] Transport Asset Protection Association. https://tapa.memberclicks.net/tapa-newsletters.

49. Manners-Bell, John. *Supply Chain Risk : Understanding Emerging Threats to Global Supply Chains.* London : Kogan Page, 2014. 9780749471101. 9780749471118..

50. Arnum, Patricia Van. *Pharmaceutical Technology.* Jan 2, 2014, Vol. 38, 1.

51. *Holograms tackle counterfeiting. European Medical Device Technology.* Lancaster.I. 5, s.l. : European Medical Device Technology, 2010, Vol. 1.

52. Sarbacker, Julie. RFID in the Pharmaceutical Supply Chain. [Online] Revenue Performance, September 12, 2005. [Cited: February 29, 2016.] http://mthink.com/article/rfid-pharmaceutical-supply-chain/.

53. Rodgers, Dirk. The case for and against RFID in pharma. [Online] August 29, 2014. [Cited: 2 29, 2016.] http://www.lifescienceslogistics.com/logistics/gdps/case-and-against-rfid-pharma.

54. Rx-360. *Rx-360 Supply Chain Security White Paper: Cargo Risk Assessment.* s.l. : Rx-360, 2012.

55. *Mechanisms of Prescription Drug Diversion Among Drug-Involved Club- and Street-Based Populations.* James A. Inciardi, Hilary L. Surratt, Steven P. Kurtz, and Theodore J. Cicero. 2, s.l. : Pain Med, 2007, Vol. 8. PMC2879025.

56. *A reassessment of trends in the medical use and abuse of opioid analgesics and implications for diversion control: 1997–2002.* Aaron M Gilson, PhD, Karen M Ryan, MA, David E Joranson, MSSW, June L Dahl, PhD. 2, s.l. : Journal of Pain and Symptom Management, 2004, Vol. 28.

57. UNODC. *World Drug Report* . s.l. : United Nations, 2012.

58. *Trends in prescription drug abuse and dependence, co-occurrence with other substance use disorders, and treatment utilization: Results from two national surveys.* Sean Esteban McCabea, James A. Cranforda, Brady T. Westb. 10, s.l. : Addictive Behaviors, 2008, Vol. 33.

59. *The relationship between past-year drinking behaviors and nonmedical use of prescription drugs: Prevalence of co-occurrence in a national sample.* Sean Esteban McCabea, , , James A. Cranforda, Carol J. Boydb. 3, s.l. : Drug and Alcohol Dependence, 2006, Vol. 84.

60. *Relationship between Nonmedical Prescription-Opioid Use and Heroin Use.* Wilson M. Compton, M.D., M.P.E., Christopher M. Jones, Pharm.D., M.P.H., and Grant T. Baldwin, Ph.D., M.P.H. s.l. : The New England Journal of Medicine, 2016.

61. UNODC. *Afghan-Drug-Survey-2009-Executive-Summary.* 2010.

62. World Bank. *Gender Statistics.* [World Bank data bank] [prod.] The World Bank. s.l. : The World Bank, 2013.

63. Sanofi-Aventis. *Drug Counterfeiting.* [Online] [Cited: 05 20, 2016.] http://ec.europa.eu/internal_market/indprop/docs/conf2008/wilfried_roge_en.pdf.

64. World Bank. World Bank. *Indicators.* [Online] http://www.worldbank.org/depweb/english/beyond/global/glossary.html#34.

65. Transperancy International. Transperancy International. *Corruption Perception Index.* [Online] 2011. http://www.transparency.org/cpi2011/in_detail.

66. UNODC. *Transnational Organized Crime in East Asia and the Pacific: A Threat Assessment.* s.l. : UNODC, 2013.

67. WIPO. World Intellectual Property Organization. [Online] http://www.wipo.int/portal/en/index.html.

68. *The Global Threat of Counterfeit Drugs: Why Industry and Governments Must Communicate the Dangers.* Robert Cockburn, Paul N Newton, E. Kyeremateng Agyarko, Dora Akunyili, and Nicholas J White. s.l. : PLOSMed, 2005.

69. *Counterfeit Drug Penetration into Global Legitimate Medicine Supply Chains: A Global Assessment.* Tim K. Mackey, Bryan A. Liang, Peter York, and Thomas Kubic. s.l. : The American Society of Tropical Medicine and Hygiene, 2015.

70. Filzmoser, M. Templ and P. *Visualization of missing values using the R-package VIM.* Institut f. Statistik u. Wahrscheinlichkeitstheorie. 2008. R package documentation.

71. Daniel Ekwall, Helmut Bruls, Daniel Wyer. Theft of pharmaceuticals during transport in Europe. *Journal of Transportation Security.* 2015.

72. *Supply Chain Management under the Threat of International Terrorism.* Yossi Sheffi, Massachusetts Institute of Technology. 2, s.l. : The International Journal of Logistics Management, 2001, Vol. 12.

73. *Rejoinder: Sure independence screening for ultrahigh dimensional feature space.* J. Fan, J. Lv. s.l. : Journal of the Royal Statistical Society Series B, 2008, Vol. 70, pp. 905-908.

74. *Regression Shrinkage and Selection via the Lasso.* Tibshirani, Robert. 1, s.l. : Journal of the Royal Statistical Society. Series B (Methodological), 1996, Vol. 58.

75. *An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests.* STROBL, C, MALLEY, J and TUTZ, G. 4, s.l. : Psychological Methods, 2009, Vol. 14. 1082-989X.

76. Breiman, Leo. Bagging predictors. *Machine Learning.* s.l. : Springer, 1996.'

77. *Random Forests.* Breiman, Leo. Berkeley : Machine Learning, 2001, Vol. 45.

78. Breiman, Leo. *Manual On Setting Up, Using, And Understanding Random Forests V3.1.* 2002.

79. O'Connell, Ann A. The Cumulative (Proportional) Odds Model for Ordinal Outcomes. *Logistic Regression Models for Ordinal Response Variables.* s.l. : Little Green Books, 2006.

80. *PARALLEL LINES ASSUMPTION IN ORDINAL LOGISTIC REGRESSION AND ANALYSIS APPROACHES* . YILDIZ, Erkan ARI and Zeki. 3, s.l. : International Interdisciplinary Journal of Scientific Research , December 2014, International Interdisciplinary Journal of Scientific Research , Vol. 1.

81. Frank E. Harell, Jr. *Regression Modeling Strategies.* New York : Springer, 2001. 2197-568X.

82. Agility Logistics. [Online] http://www.agility.com/EN/Industry-Solutions/Pages/Cross-border-trucking-service-opens-new-routes-in-Southeast-Asia.aspx#.V63AfigrLIU.

83. [Online] https://www.drugs.com/support/about.html.

84. Chubb Group of Insurance Companies. Mitigating Transit Theft. *http://www.chubb.com.* [Online] [Cited: 05 06, 2016.] http://www.chubb.com/businesses/cci/chubb1100.pdf.

85. Helmut Brüls Head of EMEA Intelligence, FreightWatch International, Brussels and Daniel Wyer EMEA Intelligence Analyst, FreightWatch International, Dublin. *Pharmaceutical Cargo Theft in Europe: A REALISTIC VIEW OF THE CURRENT TRENDS, CHALLENGES, AND FINANCIAL IMPACTS.*

86. World Health Organization. *General information on counterfeit medicines.* s.l. : WHO, 2006.

87. *Counterfeit Drug Penetration into Global Legitimate Medicine Supply Chains: A Global Assessment.* Tim K. Mackey*, Bryan A. Liang, Peter York and Thomas Kubic. 6, s.l. : The Americal Journal of Tropical Medicine and Hygiene, 2015, Vol. 92.

88. Jeremy M. Wilson, Roy Fenoff. *The Health and Economic Effects of Counterfeit Pharmaceuticals in Africa.* s.l. : Michigan State University Anti-Counterfeiting and Product Protection Program (A-CAPPP), 2011.

89. Worldfreightrates. [Online] [Cited: 10 29, 2016.] http://www.worldfreightrates.com/.

90. Eötvös Loránd University. Graph Algorithms. [Online] http://compalg.inf.elte.hu/~tony/Oktatas/TDK/FINAL/Chap%2013.PDF.

Appendix

1. Code for linear regression

Create train and test datasets

```
set.seed(683)
smp_size <- floor(0.80 * nrow(diversion1))
train_ind <- sample(seq_len(nrow(diversion1)), size = smp_size)
train <- diversion1[train_ind, ]
```

Create design matrix

```
design <- A ~ .^2
Z2 <- matrix(as.numeric(model.matrix(design,data=test)),nrow=13)
X <- matrix(as.numeric(model.matrix(design,data=train)),nrow=80)
trainX <- as.matrix(train[-1])
test <- diversion1[-train_ind, ]
test1 <- as.matrix(test[1])
```

SIS and ridge regression

```
for(j in 1:nq){
        r     <- abs(cor(trainY,X))
        keep   <- which(rank(-r) <= q[j])
        ridge <- cv.glmnet(trainX[,keep],trainY[,keep],alpha = 0,standardize=T)
        MSE1[j] <- mean((trainY-predict(ridge,trainX[,keep]))^2)
}
```

SIS and Lasso regression

```
for(j in 1:nq){
        r     <- abs(cor(trainY,X))
        keep   <- which(rank(-r) <= q[j])
        lasso <- cv.glmnet(trainX[,keep],trainY[,keep],alpha = 1,standardize=T)
        MSE2[j] <- mean((trainY-predict(lasso,trainX[,keep]))^2)
}
```

SIS and ordinary logistic regression

```
VOLSresp <- data.matrix(train[1])
resp <- data.matrix(scale((train[1]),center=FALSE))
q    <- seq(10,80,2)
nq   <- length(q)
MSE <- matrix(0,nq)
MSE3 <- matrix(0,nq)
set.seed(683)
# 5 fold cross validation with replacement
fold <- sample(1:5,80,replace=TRUE)
table(fold)
n <- 80
for(j in 1:nq){

            Yhat1 <- rep(0,n)
            Yhat3 <- rep(0,n)
            for(f in 1:5){

              Ytrain <- VOLSresp[fold!=f]
              Ytest <- VOLSresp[fold==f]

              Xtrain <- data.matrix(X[fold!=f,])
              Xtest  <- data.matrix(X[fold==f,])

              r      <- abs(cor(Ytrain,Xtrain))
              keep   <- which(rank(-r) <= q[j])

              Xtrain[,keep] <- data.matrix(Xtrain[,keep])

              ols <- lm(Ytrain ~ (Xtrain[,keep]))
              ols1    <- ols$coef
              ols1[is.na(ols1)] <- 0
              yh      <- ols1[1] + Xtest[,keep]%*%ols1[-1]
              Yhat3[fold==f] <- yh
            }

            SE3 <- (resp-Yhat3)^2
            SE3[is.na(SE3)] <- 0
            MSE3[j]<- mean(SE3)
}
```

2. Code for nonlinear regression

Random forest

```
library(randomForest)
library(caret)
train.rf.imp <- rfImpute(Diversion ~ ., data=train.rf)
set.seed(50)
ranfor.imp <- randomForest(Diversion ~
.,data=train.rf.imp,ntree=2000,mtry=3,importance=T)
j <- which.min(ranfor.imp$mse)
i <- ranfor.imp$mtry
set.seed(50)
ranfor.imp.min <- randomForest(Diversion ~
.,data=train.rf.imp,ntree=j,mtry=i,importance=T)
print(ranfor.imp.min)
# Generate Plots
plot(ranfor.imp.min)
varImpPlot(ranfor.imp.min)
importance(ranfor.imp.min,type=1)
VI_Train <- as.vector(varImp(ranfor.imp.min))
```

3. Code for Ordinal logistic regression

Create test and train datasets

```
smp_size <- floor(0.75 * nrow(C1))
set.seed(50)
train_ind <- sample(seq_len(nrow(C1)), size = smp_size)

train <- C1[train_ind, ]
test <- C1[-train_ind, ]
test1 <- test[-1]

smp_size2 <- floor(0.75 * nrow(C2))
set.seed(50)
train_ind2 <- sample(seq_len(nrow(C2)), size = smp_size)

train2 <- C2[train_ind2, ]
test2 <- C2[-train_ind2, ]
```

Ordinal Logistic Regression

```
library(nnet)
require(MASS)
model <- multinom(CII ~.,data=train)
summary(model)
predict(model,test)
confmat <- table(predict(model),train$CII)
print(confmat)
confmat3 <- table(predict(model,test),test$CII)
print(confmat3)
predclasserror <- 1-sum(diag(confmat3))/sum(confmat3)
logLik(model)
confint(model,level=0.95)
require(effects)
model.eff <- Effect("train2",model)
```