EMPIRICAL AND SIMULATION BASED ANALYSIS OF DETAILING ACTIVITIES OF A PHARMACEUTICAL COMPANY IN TURKEY

NURİ SEZAY DEMİRBACAK

BOĞAZİÇİ UNIVERSITY

2014

EMPIRICAL AND SIMULATION BASED ANALYSIS OF DETAILING ACTIVITIES OF A PHARMACEUTICAL COMPANY IN TURKEY

This thesis is submitted to the

Institute for Graduate Studies in the Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Management Information Systems

by

Nuri Sezay Demirbacak

Boğaziçi University

2014

Empirical and Simulation Based Analysis of Detailing Activities of a Pharmaceutical Company in Turkey

The thesis of Nuri Sezay Demirbacak

has been approved by:

Assoc. Prof. Osman Darcan ______________________

(Thesis Advisor)

Assoc. Prof. Bertan Badur ______________________

Assist. Prof. Zeynep Ata ______________________

August 2014

ABSTRACT

Nuri Sezay Demirbacak, "Empirical and Simulation Based Analysis of Detailing Activities of a Pharmaceutical Company in Turkey"

In this thesis, the impact of detailing activities on prescription behaviour of physicians and their carryover effects are studied using the geographic location based sales and detailing activity data of five products from a leading pharmaceutical company in Turkey. First, the relationship of sales to detailing activities and to activities made in different customer segments are analyzed using a linear regression model. It has been found that a significant relationship exists between total detailing activities and sales for four out of five products. As to the activities made in different customer segments, significant results are observed for only one product. Secondly, the outcomes of the regression are used to develop an agent based model to predict the impact and the carryover effect of detailing activities on different customer segments. The outcomes of the simulation model, when run with the best fitted impact and carryover ratios for the selected regions, reveal 10-15% deviation amongst the actual monthly values and the predicted values.

ÖZET

Nuri Sezay Demirbacak, "Türkiye'deki bir İlaç Firmasının Tanıtım Faaliyetlerinin Veri ve Benzetim Bazlı Analizi"

Bu tezde, Türkiye'de önde gelen bir ilaç firmasının beş ürününe ait, bölgesel bazlı satış ve ilaç tanıtım verisi kullanılarak, tanıtım faaliyetlerinin doktorların reçete yazma eğilimleri üzerindeki etkisi ve akılda kalınırlığı incelenmiştir. İlk olarak, toplam tanıtım faaliyetlerinin ve müşteri segmentleri bazında yapılan faaliyetlerin, satış üzerindeki etkileri, doğrusal bağlanım modeli kullanılarak analiz edilmiştir. Bu analiz sonucunda, beş ürünün dördünde satış ve toplam ilaç tanıtım faaliyetleri arasındaki bağıntı istatistiki olarak anlamlı bulunmuştur. Müşteri segmentleri bazında incelendiğinde ise bir ürün için anlamlı sonuçlar elde edilmiştir. İkinci aşamada, anlamlı sonuçlar bulunan ürünün değerleri kullanılarak, tanıtım faaliyetlerinin farklı müşteri segmentleri üzerindeki etkisi ve akılda kalma oranını hesaplayan, ajan bazlı bir benzetim modeli geliştirilmiştir. Benzetim modeli, hesaplanan en uygun etki ve akılda kalma oranları ile seçilen bölgelerde çalıştırıldığında, gerçek aylık pazar payı değerleri ve benzetim sonucu elde edilen değerler arasında %10-15 sapma gözlemlenmiştir.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Osman Darcan who has provided endless support from the start to the end of this study. His consultancy and friendship has helped me stay motivated at all times, while his flexibility and understanding during discussions has given the possibility to explore new ways and ideas.

I would like to thank professor Bertan Badur for all his inspirational ideas and for helping me find my way when I got lost in the vastness of statistical analysis and models. His knowledge in the field coupled with his sincere enthusiasm has helped greatly in achieving the results and finding the methodologies contained in this thesis.

I would like to thank Ümit Aydın and Ali Orhon, who supported me at all times and allowed me to use the data analyzed in this study. Without them, this study wouldn’t have been possible.

I want to thank Zeynep Ata for guiding me in putting together the pieces in this thesis in a meaningful analysis and for her enthusiasm in the subject of the thesis.

I want to thank Göker Erdoğan, Hakan Karşılar, Volkan Dündar and Yalçın Akın Duyan, for all the friendship, academic knowledge and motivation they have provided, on demand, anytime, anywhere.

Lastly, I want to express my biggest thanks and gratitude to the biggest contributor and the woman behind the scenes; Esra Aynacı. The type and depth of support she has provided is unmatched to anything I have ever experienced before.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Pharmaceutical industry has depended primarily on its sales representatives for driving sales. There are approximately six million sales people in the United States, costing a trillion dollars every year to their companies who employ them (Zoltners, 2005). It is estimated that seven billion dollars of this one trillion is spent by the pharmaceutical industry alone (Manchanda et al., 2005). The primary objective of the sales representative, working in a pharmaceutical company; is to visit customers and promote products. The optimization of this process has given birth to many different study areas, i.e. sales territory alignment, customer classification, sales territory optimization and call planning.

While many have focused on the optimization of sales territories and field force sizing, there is little research on the modeling of physician's behavior of prescription, which may be derived from the statistical analysis of field activity and sales. This thesis is made available by a sponsoring pharmaceutical company, who have agreed to supply its field force activity data of 5 products for the Turkey market alongside with its and its competitors sales figures. The data is first analyzed using linear statistical analysis for finding correlation between the sales in a geographic region and detailing activity done in that region for different customer segments, then an attempt is made to construct an agent based model, based on the parameters achieved from the statistical analysis, to simulate the effect of detailing activities on sales and predict a plausible carry over effect. Amongst the relatively few articles in academic literature, the ones that are of most importance and relevant are; study of correlation of activity and sales of pharmaceutical companies (Taneja, 2008), simulation of carry over effects using neural networks (Yi, “Anand” Anandalingam, & Sorrell, 2003) and optimization algorithm for maximizing return on investment for detailing activities (Agnetis, Messina, & Pranzo, 2010).

This thesis will use similar concepts of approaching to the same problem but will use different methodologies and tools to achieve the results. We will first develop a statistical model and analyze the correlation of detailing activities and sales in a region for each product, from which we will derive parameters to be used to build a simulation of real life prescription behaviors and the carry over effects for different physician segments related to detailing activities for a particular product in a single region.

This thesis presents the study in the next three chapters; Chapter 2 includes the background information needed to understand the pharmaceutical environment and an extensive review of the literature with the problem definition, Chapter 3 presents the statistical models and correlation test results, and Chapter 4 puts forward the details of the agent based model constructed, its outputs and discusses its possible uses, while ending with a conclusion of the thesis with possible future enhancement suggestions.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

In this chapter, a solid background for understanding the problem definition this thesis deals with is provided, while covering an extensive part of the literature, exploring related works. It will also cover some background on the tools and approaches used in this study.

Sales Coverage Unit (SCU) Definitions in Pharmaceutical Industry

The sales territory alignment problem may be viewed as the problem of grouping small geographic sales coverage units (SCUs) into larger geographic clusters called sales territories in a way that the sales territories are optimal according to managerially relevant alignment criteria (Zoltners & Sinha, 1983).Pharmaceutical companies base their SCUs on what is widely called as bricks, which are the highest granularity of geographic regions by which the local sales data provider collects and provides sales information to pharmaceutical companies. Such sales data provider companies gather information from wholesalers to determine the sales of all drugs in the detail of their dosage and packaging made to all pharmacies. The data provider then maps the geographic location of each pharmacy to a defined set of geographic regions called bricks (SCUs) and provides the data to pharmaceutical companies by masking the sales data with the brick information. The data provided includes sales in units and in local currency, for a particular month, in each brick, for every form of drugs (dosage & packaging), accompanied by information such as the owner, distributor, molecule, ATC classes and other useful information of each drug. ATC Class (Anatomical Therapeutic Chemical (ATC) Classification System) is particularly important because each product has one ATC Class which defines its active curing disease area and hence all products with the same therapeutic class define the products and its competitors, what we will call as the market of a product. The market share and the performance of the drug will be calculated within the market which the product belongs to.

Since like all commercial companies, the pharmaceutical industry bases their resource and field force allocation on identifiable bricks (Lodish, 1975), for the sake of efficiency and simplicity, the companies use the defined set of bricks for forming and aligning their territories, subsequently assigning human resources to each defined groupings of bricks, usually called sales territories. Each sales territory then is grouped into districts, based on the complexity and the number of bricks covered by the sales force. A typical sales hierarchy in a pharmaceutical sales organization may be as shown in Figure 1**Error! Not a valid bookmark self-reference.**.

Brick → Territory → District → Sales Team → Local Subsidiary

Figure 1. The levels from bottom to upper levels of the sales hierarchy for a typical pharmaceutical company

Product and Portfolio Assignment to Field Force Sales Teams

Pharmaceutical companies define which set of products they will promote using their people sales force. The channel of sales representatives is a strong and primary influence of promoting drugs and their effective symptoms (Zoltners, 2005). A sales representative visits or in business terms *calls upon* a physician or pharmacist to convey a message related to a product. Since the number of resources is finite in the field force, companies wish to allocate these resources to the products which derive strong return on investment (ROI) or have high potential for growth (i.e. new product launches) (Bates, Bailey, & Rajyaguru, 2002). As an example, imagine a field force consisting of 300 sales representatives and suppose the company wants to promote 30 products. Each product has a potential target audience; usually depending on the ability of the health care professional (HCP) to influence the sales of the product, either by direct influence; like prescribing the product or approving purchase of drugs to a medical institution, or by indirect influence; where an HCP may be at a teaching or training position at an institution or the person in question may be a pharmacist who may have the power to advise the customer to use the product – usually as a substitute to another competitor product prescribed by the physician. Since the 30 products would have different target audiences, the efficient way is to group them to what is called *product portfolios* and assign different groups of people called sales teams to these portfolios (Lodish, 1975). So, in other words, the 300 sales people should be grouped into teams of people with a specific product portfolio, which would consist of usually more than one product. The company's sales force effectiveness department will decide in the optimum assignment of product portfolios and how many people would work in those teams. In this example, it may be possible that there are 8 teams formed, with different number of products assigned to each team and 300 people are divided into smaller groups to be assigned to these teams. This way a person in a single team will be able to convey messages of products to an HCP in a single visit as the team structure will allow for him to have a target audience grouped depending on his product portfolio. In doing so, the pharmaceutical companies not only optimize the resources spent during a visit to a single HCP for promoting similar products but also arrange their other business lines, like marketing and operations, to better serve a specific group of customers with similar traits and needs.

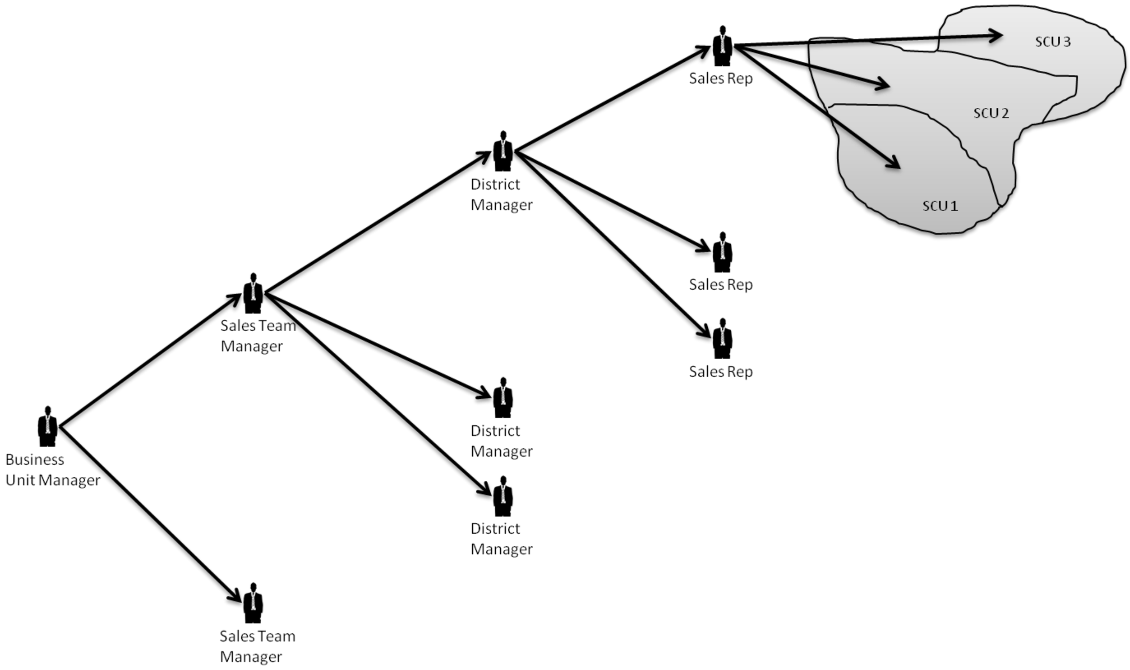

Figure 2. Illustration of the sales organization hierarchy in a typical pharmaceutical company

Sales teams are basis of product promotions in a pharmaceutical company's sales force. Each team acts as an individual sales force, with their defined set of products, target HCPs, budget, sales targets and executive sales management. The teams also have a different sales territory alignment, meaning each team constructs their own groupings of bricks, territories and districts, the hierarchy may be observed in Figure 2. From right to left, the sales representatives are assigned bricks and they report to the district manager, thus the total of those bricks of the sales representatives who report to the same district manager are the bricks which form the whole district. Same applies for the sales team manager to whom all district managers report and the summed area of the bricks of each sales representative in each district will determine the overall geographic coverage of the team. It is also common for companies to produce internal profit and loss analysis based on the sales team performance.

Sometimes, due to a large number of products with similar target audiences, the products which share a common HCP list need also be broken down into smaller teams; usually if the products demand very specific actions to be taken or if the size of the field force exceeds the people management capacity of sales managers (i.e. district managers who manage sales representatives)

The Customer Plateau of Pharmaceutical Companies

Pharmaceutical companies have, due to regulations and the structure of the market, many parties which may have traits of customers. A customer (sometimes known as a client, buyer, or purchaser) is the recipient of a good, service, product, or idea, obtained from a seller, vendor, or supplier for a monetary or other valuable consideration (Blythe, 2008). For pharmaceutical companies, this definition does not necessarily point out to a defined group of people or entity. The good in question, that the company sells and promotes, are drugs. Drugs are prescribed by physicians, which are supplied to wholesalers, who in turn sell to pharmacies from where the patients – the actual consumers of the product – buy the drug. The physicians prescribe the drugs but the patient buys the products from a pharmacy. The payment of the product is not solely done by the patient; it is usually the health insurance provider of the patient (Kotler, 2000). For a country like Turkey, the main provider is the government for 90% of the patients (Tokgöz, 2010).The customer for a pharmaceutical company is divided into 4 main groups which may be summarized as:

- Patients – who receive and use the product.

- Health Care Professionals (HCP): Physicians – who prescribe the drug and Pharmacies – who may influence which drugs will be used by the patient.
- Payers – usually the health insurance providers or the government.
- Providers – which are usually wholesalers who act as an intermediary means in the supply of the product between the drug manufacturer and the points of sales.

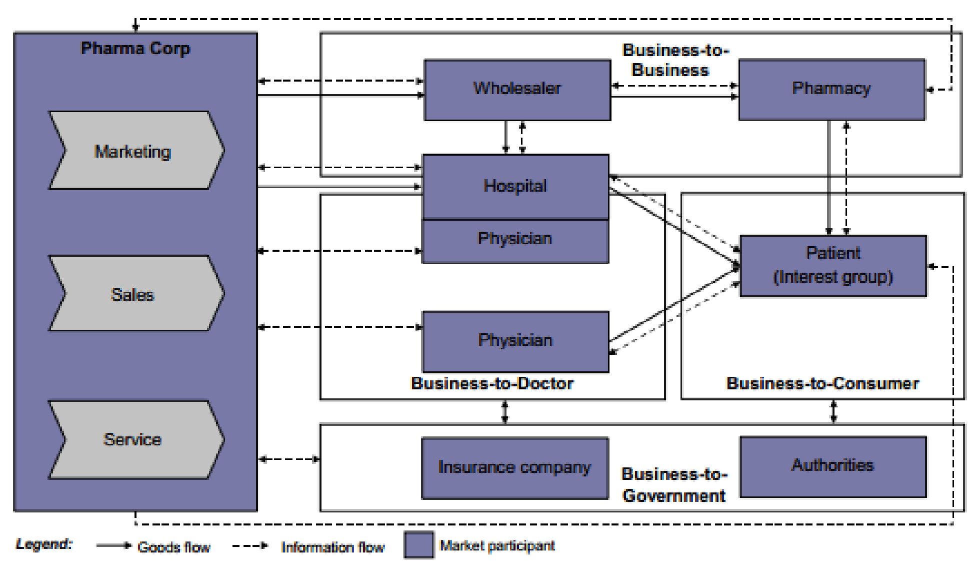

Figure 3. The stakeholders and their interactions in a typical pharmaceutical sales chain

The interaction between the four groups are shown in Figure 3.This diversification in the customer plateau of pharmaceutical companies has usually resulted in companies to have different business groups addressing the needs of different customer groups. Customer relations or trade departments deal with providers, accounts receivable or account managers manage payers and payments, the marketing team target patients for drug and treatment awareness raising, loyalty or compliance campaigns while the main sales team – the field force – focuses on physicians and pharmacies. In most cases, human resources are designated for the HCPs; mainly because they are the primary source of initiating and determining the sales of products. This thesis will focus on the sales teams whose primary focus ison the HCPs and who devote their time, energy and resources to this target audience. Throughout the thesis, the usage of the word customer will refer to HCPs, if not explicitly stated otherwise.

Customer Data in Pharmaceutical Companies

The customer data quality and accuracy is one of the most crucial assets for a pharmaceutical company, as they focus most of their human and monetary resources to customers. The driving factor in sales is found to be the promotions of products made to customers, excluding the success of the treatment of the symptoms which a drug is developed for – there are various studies in literature which suggest existence of a direct correlation between calls and promotional one-on-one activities with the customer and sales (Agnetis et al., 2010; Al-Hamdi, Hassali, & Ibrahim, 2012; Gönül, Carter, Petrova, & Srinivasan, 2001; Lerer, 2002; Lodish, 1975; Taneja, 2008; Zoltners & Sinha, 1983).Thus, in several countries, there are corporations who acquire the information of HCPs, their names, specialties, where they work and other useful attributes, and these corporations sell this information to pharmaceutical companies. The information includes the health care organizations like hospitals, departments, clinics, pharmacies, group practices, private clinics as well as HCP individual information, such as name, surname, specialty, working relations with organizations. This information also identifies the address of all organizations and the individuals aligned with the bricks. This way, the pharmaceutical companies can match which HCPs are in a particular brick and relate it to the sales of their own or competitor's products in that brick. The companies cannot, however, determine which particular individuals or organizations are related to sales of products, unless the regulations allow for it. Currently only few countries have such regulations which permit the distribution of such information (i.e. USA) (Taneja, 2008; Yi, 2008). In Turkey, a pharmaceutical company can only know which HCPs and organizations exist in a brick and know the sales in a particular brick, but cannot relate a specific sales figure to any of the individual entities. In this thesis we will build on the assumption that the sales made in a particular brick is derived from the customers within that brick. There some exceptions to the case, such as a patient taking a prescription from a hospital but going outside the brick to get the drugs from the pharmacy, but these cases are rare and may be neglected.

Customer Segmentation in Pharmaceutical Companies

Once the sales territories, teams and their product portfolios are determined at the start of each fiscal year, the employees are assigned to sales territories and are distributed sales targets based on their bricks. This activity consists of 3 main processes:

1. Sales force sizing; where the number of employees are determined which will promote each product or product portfolio (Baier, Carballo, & Chang, 2012)
2. Territory alignment; by which bricks are grouped to form sales territories
3. Sales force location; in which the home base for each sales territory and representatives are determined (Drexl, A.; Haase, 1999)

Once each employee is assigned to a territory and hence to a sales team, they will have a customer list depending on the properties selected for that sales team and where they cover as sales territory. A *customer list* emerges for all sales representatives, which include the HCPs who are located in the covered bricks of that sales territory and that match the specialty or organization type set for that sales team. Usually SFA (Sales Force Automation) or CRM (Customer Relationship Management) systems automatically create and maintain such lists as the HCPs may move from one geographic location or organization to another during the course of time (Baier et al., 2012; Puschmann & Alt, 2001). The sales representative then starts planning and making calls to his/her customer list, everyday visiting different customers to increase and support the sales targets for their sales territory, for the products of their sales teams which they are assigned with.

In a typical sales territory, there are usually more customers than a sales representative can visit in a single month, thus it becomes crucial for which particular customers the field force consumes their precious time to visit and promote drugs. To overcome this prioritization issue in call planning, the marketing departments of pharmaceutical companies have developed customer segmentation techniques which enable sales people to differentiate between the most effective and the least effective customers. The segmentation of customers are revised, usually, on annual basis and define if the influence of the customer in sales of a particular product is high or low and if the customer yields any potential in becoming an important sales driver (Lerer, 2002).

The segmentation of customers in pharmaceutical industries commonly use a two criterion based matrix, where one axis marks the sales potential or the number of prescriptions a physician makes and on the other axis marks the loyalty of the physician for a specific product – which is the percentage of products of the company prescribed by the physician. This division into quadrants may be seen in Figure 4.

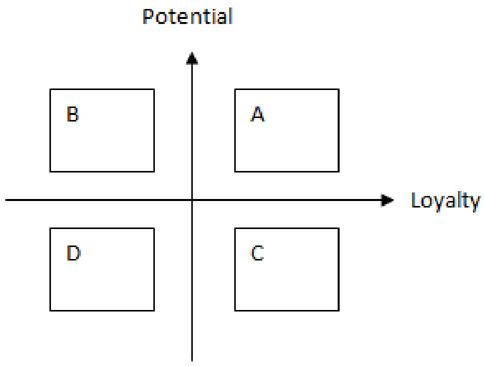

Figure 4. The segmentation matrix and segment names of a typical pharmaceutical company

Figure 4 shows the y axis as the sales potential of the customer, usually measured by how many drugs a physician prescribes for the particular symptom or disease, in other words the total sales derived from that single physician within the selected symptoms. The x axis shows the loyalty of the customer, usually measured by the percentage of how much prescriptions are made for the company's products amongst the total number of products prescribed, captured in the y axis. This segmentation per each product is then displayed in the SFA or CRM tool for the sales representatives to know which customers they should visit most, the usual importance of ordering being A, B, C, D segmented customers respectively. Usually each segment of customers are assigned with a target frequency, which suggest a particular number of times a physician belonging to that segment should be visited during a sales cycle. Each segment may be summarized as:

- A Customers: these are the customers which bring the biggest market share to the company, meaning they prescribe more than average quantities of products and are tentative to prescribe the companies' products compared to competitors' products.

- B Customers: these are the customers which have a high sales potential, but are tentative to prescribe the competitors' products compared to companies' products.
- C Customers: these are the customers which have a low sales potential, but are tentative to prescribe the companies' products compared to competitors' products.
- D Customers: these are the customers which have a low sales potential, and are also tentative to prescribe the competitors' products compared to companies' products.
- Not Segmented Customers: these are the customers where there's no data available for their prescription amounts or the loyalty to the pharmaceutical company.

After the territory alignment activities and segmentation is finalized, the sales representatives are then expected to maximize the number of customers they visit while trying to realize number of frequency set for each customer segment. This process of actual call planning and conducting, are in most cases, not strictly handled and is usually autonomous, where the representatives and the district managers will decide on which customers should be visited in the given limited time of the sales representative within a single month. The companies usually depend on the experience and decision making quality of the sales representatives and district managers for this process.

Agent Based Modeling

This thesis will utilize agent based modeling (ABM) for developing the simulation model to predict the memory and carryover effect of customers. MC Macal and MJ

North has an excellent article putting forward the importance and differentiating factors of ABM (Macal & North, 2010):

> Agent-based modelling and simulation (ABMS) is a relatively new approach to modelling complex systems composed of interacting, autonomous ‘agents’. Agents have behaviours, often described by simple rules, and interactions with other agents, which in turn influence their behaviours. By modelling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviours can be observed as it gives rise to the behaviour of the system as a whole. By modelling systems from the ‘ground up’—agent- by-agent and interaction-by-interaction—self-organization can often be observed in such models. Patterns, structures, and behaviours emerge that were not explicitly programmed into the models, but arise through the agent interactions. The emphasis on modelling the heterogeneity of agents across a population and the emergence of self-organization are two of the distinguishing features of agent-based simulation as compared to other simulation techniques such as discrete- event simulation and system dynamics. Agent-based model- ling offers a way to model social systems that are composed of agents who interact with and influence each other, learn from their experiences, and adapt their behaviours so they are better suited to their environment.

The approach of agents fit well with scenarios like sales force activities where one can observe distinct attributes of each agent while being able to categorize these attributes and define relationships and interactions amongst them. This thesis will primarily use the ABM to simulate a model that best fits day-to-day visits and their impacts on physicians of sales representatives. The model will also help predict the best fitting values to define the carryover effect.

Relevant Studies in Literature

The academic literature includes solutions to many problems and optimization needs in the field pharmaceutical sales force allocation, i.e. sales territory alignment (Zoltners, 2005), customer segmentation (Lerer, 2002), and sales territory optimization (Baier et al., 2012).

Sales territory alignment has been the primary focus of the academia from the early days. Some important models proposed include single product, single period, single territory (Lodish, 1975) and multiple products, multiple bricks and territory (Central, 1998; Ihaka, R and Gentleman, 1996) optimization models. The former article is considered to be a milestone article which puts out a mathematical model based on a heuristic algorithm employing a linear optimization in building sales territories and optimizing the territory sizes for efficiency. The latter articles build upon the model proposed by Lodish and take it further by enabling several products and territories to be used, while trying to employ other mathematical models such as mixed integer programming. These articles only focus on the best formation of sales territories and do not consider any real life visit scenarios or their relation to sales.

Typically, articles on call planning and physician level prescription behaviour analysis have mainly tried to address the optimization of a sales response function which represents the total ROI expected from a sales effort within a given time period (Agnetis et al., 2010). Agnetis et al, have developed an S shaped heuristic sales response function for each different segment of customers in Italy market for a pharmaceutical company. They then ran an optimization method to find the best distribution of sales force calls to 46,000 physicians across the country. Their study suggests up to 10% of increase in efficiency in terms of ROI. The model they have used is also suggested to be useful for running scenarios with different visit patterns to see the effects in sales although their work did not include such an application of the model. The Italy market is similar to Turkey market where physician level prescription information is not available and so they have to be analyzed at the brick level.

Gönül has done the most relevant work to the subject covered in this thesis, as she has focused on the carry over effect of detailing activities made to physicians and the change in their behavior when a detailing activity is done. In one of her work, she develops a multinomial logit model, to understand the effect of detailing activities, price of the drug and insurance provider on physicians' prescription behaviour (Gönül et al., 2001). Gönül concludes the study with mentioning that brand loyalty cannot be fully observed for physicians; she has found that only 2% of physicians in the study could be defined as loyal to a brand, implying that detailing visits are highly influential and are crucial for sustaining market share. She also argues the complexity of effects and the improbability of accurate cause-effect measurements when it comes to physician prescription behaviour. Their study develops a memory decay effect on a monthly basis to match with market share as the response variable. Gönül and colleagues compute a model that fits for all products they have considered, this is enabled by the availability of physician level prescription data in US and the model is based on statistical analysis using nonlinear correlation of physician prescription data versus price, visits, samples given and the patients' health care provider.

John Yi has calculated the carry over effect at the physician level for a pharmaceutical company in the US. He has used 2 years' worth of data, aggregated at the quarterly level and has included 10 segments of customers. Yi has utilized neural networks in building his model and used nonlinear models to understand the carry over effect of detailing activities made to physicians. He has excluded other promotional activities due to concerns about the correctness of the data. He was able to attain carry over effects and responsiveness levels of physicians to detailing activities at the individual level, all of which he then used to acquire carry over effects for each customer segment. Yi has conducted this study only for a single product, and has found a carryover effect of 0.73 as an average per quarter. Meaning that, the effect of a detailing visit as depicted with 1, would drop to 0.73 as an indication of 27% drop in units prescribed by the physicians. (Yi, 2008)

Other studies, on understanding the effect of detailing activities on physicians' prescription behavior and its influence power, are mainly conducted in markets like United States, where prescription data is available at the physician level (Gönül et al., 2001; Yi et al., 2003). While others, in other markets similar to Turkey, did not have access to direct data or had restricted access, or used survey methodologies to collect data, which are all restricted ways used when compared to what this thesis will use (Al-Hamdi et al., 2012; Taneja, 2008; Wright & Lundstrom, 2004).

While some of these models try to maximize the return on investment given in a time period or horizon and come with a model to suggest best possible distribution of limited call hours to a limited target customer list, the latter examples given are built to understand the carry over effect on customers. This thesis will focus on the latter study area, focusing on understanding the carry over effect in the Turkish market; where physician level prescription information is not available and sales figures are only available at brick level.

CHAPTER 3

EMPIRICAL MARKET SHARE MODEL

This chapter will start with describing the sample data used in this study and the methodologies by which the data was attained. All data is courtesy of an international pharmaceutical company who will be kept anonymous. The data used is for the Turkey market of this company and the data will be masked for security and copyright purposes. The company will be called T-Pharma co. for the rest of this thesis, which is an imaginary name devised for the purpose of anonymity.

Problem Statement

We will aim first to explore the correlation between sales, measured by market share and detailing activity, measured by call count, using a linear regression model, to understand the level of predictability of sales as function of calls made to different segments of customers for each product. We will use R software for statistical analysis purposes, which has been become a standard for running linear regression models on panel data in the academic world (Ihaka, R and Gentleman, 1996). After finding a suitable linear regression model, we will then utilize the attained outputs and correlation factors in building an agent based simulation, by which we will try to find the best suitable carryover effects within a single brick for a single product. NetLogo will be used for the agent based simulation development, which is an agent based modeling tool, used in wide area of research for simulation purposes (Wilensky, 2013).

Sample Data

There are mainly 3 sets of data used in this thesis:

1. Segmentation Surveys and outputs: This data is a collection of surveys conducted by the company, filled out by its field force employees to determine the sales potential (how many units of drugs are prescribed by physicians) and the loyalty (how many of those units of drugs belong to the company in question) of customers. The outcome of the surveys is the segmentation (A, B, C, D) of the customers which are available on product level per customer.
2. Detailing activity and calls: This is the data extracted from the CRM system of the company, which captures each promotional activity done to customers with the information of which products were detailed. The calls are aggregated at month and customer segmentation level.
3. Sales: This is the sales of the company and its competitors' products for a 6 month time period at brick and month level.

The sample data consists of monthly aggregated sales and customer visit data for a 6 month period between March and August 2013. The data is selected for a stable time period in the market; since year beginnings and ends tend to fluctuate according to stock building activities of pharmacies and wholesalers during these time intervals. The data includes only customer visit data for T-Pharma Company and sales data for all the pharmaceutical companies operating in Turkey in the selected time period.

The sample data is unique for every product at a given month for a given brick. There are 1001 bricks in Turkey but only bricks with customer visits more than 5 are selected for each product, since the bricks where there is less than 5 customer visits, there tends to be of little importance of sales in that region, causing small variances in sales to result in significant fluctuations and unstable data trends. The sample data size is 12,650 rows of information for six months for five products. The activities made by 278 sales representatives are included in the analysis. Each product has approximately 2400 rows of data for the six months in about 700-900 bricks.

The sales data is retrieved from a sales data provider in Turkey, which is the sole company providing sales information of wholesaler to pharmacy for each product. The accuracy of the data is estimated to be 98% with the actual total sales in Turkey (Tokgöz, 2010). The data consists of sales of every package of drug sold in Turkey (i.e. Asprin 50mg 20 tablets) for each calendar month. The sales are in standardized packaging units which is the normalized form of packaging units data, to calculate a meaningful market share. A small sample of the sales data may be seen in Table 1.

Table 1. Sample of the Sales Data.

ATC Class	Product Brand	Company	Month	Units	Market Share
ATC Breast Cancer	Brand X	T-Pharma	2013 - Feb	50	1.3%
ATC Breast Cancer	Brand A	Competitor	2013 - Feb	150	4.2%
ATC Breast Cancer	Brand Y	Competitor	2013 - Feb	54	1.4%
ATC Breast Cancer	Brand Z	Competitor	2013 - Feb	23	0.7%

Sales Data & Calculations

The sales base units are measured in standardized packages sold for a given period of time for a single product. The sales performance of a product will be measured with the below KPI (key performance indicator):

- Market share (%): this is measured by the relative percentage of a product sales compared to the total market. Each product belongs to a market by its ATC class and all other drugs belonging to the same class are the competitors. The market share of a product is the percentage of sales in a given period and brick compared to the total sales of all products of the same ATC class.

The sales data is the sales of products from wholesalers to the pharmacies. This feature of the data yields a limit in understanding the direct effect in sales of the detailing activities made to physicians. The distribution and supply chain of the pharmaceutical companies in Turkey are all same due to government regulations. The steps of sales of products are as shown in Figure 5.

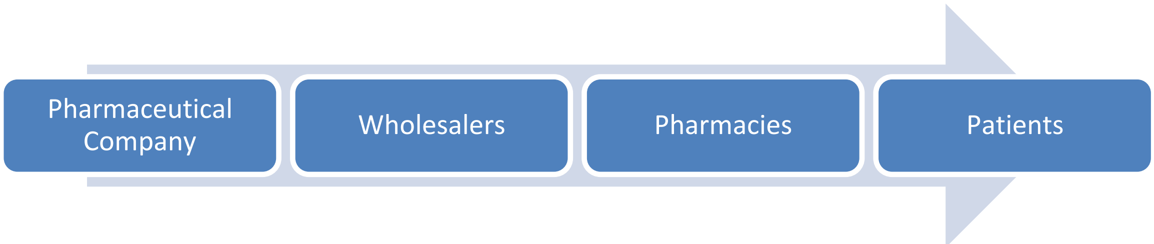

Figure 5. The supply chain of a drug in Turkey

The sales data used in this analysis is the sales from wholesalers to pharmacies while the main effect of customer visits is expected to result in increase in volume of prescriptions for a single product, which would be observed directly in the last step in the supply chain which is pharmacies to patients. For observing the effect of customer visits to the sales from wholesalers to pharmacies, the process is to wait for the pharmacy to run low on stocks for that product and request supply for that product from the wholesalers. In some cases, pharmacies may order new products from the wholesaler due to campaigns, or future price increase expectation or simply to build more stock. This makes it harder to understand the effect of visits made to customers, but we will use this data since it is the only sales data with competitor information available in Turkey. Due to this latency of observation in sales response, we will have to model our regressions by matching sales with each previous month separately to understand the true effect of customer visits in a single period to the sales of that product.

Customer Visit – Field Force Activity Data

The customer visit data is in the same form as the sales data, based on bricks and months for each individual product being analyzed. A customer visit in pharmaceutical industry means the promotion of a product to customers, which is expected to result in an increase in prescription of that drug if the detailing of the product is effective and is considered successful. Customers are visited by field force representatives and each HCP is visited by a single sales representative for a single product, which is due to optimization of sales resources by pharmaceutical companies and the need to decrease disruption and provide a single point of contact for the customer.

As previously detailed in Chapter 2, the customers have a different importance in terms of each product and are classified into groups of 5; A, B, C, D and No-Segment customers.

Activities may have a different range of content, ranging from detailing the strengths and positive clinical trials of a product to other marketing content such as campaigns or product messages. But all activities have the same objective; which is to increase the sales of a product by gaining a competitive advantage from the perspective of the customer compared to the competitor products in the same ATC class. The sample data for field force activity is within a timeframe of nine months, because for each month of sales data, previous three months of activity data is analyzed. Below are the KPI which may be used in measuring the activity performance of a sales representative:

- Call count: this is the scalar value which measures the frequency of visit, each visit counting as 1 for a single product. This is also measured for each customer segment separately:
    - Call Count for A Customers, B Customers, C Customers, D Customers, and No-Segment Customers

Hypothesis and Model

The aim of the tests will be to find a correlation between the sales of a product and the detailing activities done for that product in one of the three previous months. After analyzing the correlation of sales and activity, and after finding the lag between sales and activity in Hypothesis 1, a more detailed analysis will be done for understanding the contribution of visits made to different customer segments to sales in Hypothesis 2. The product(s) which will reveal solid results for Hypothesis 2 will be the basis for our parameters and assumptions in building the model for our agent based simulation.

Because of the sales cycle of the product, as previously shown, there is a lag between the actual changes in the prescription which are influenced by the detailing activities and the sales data reflecting these patterns. This lag is highly dependent on the nature of the sales cycle of the product as well as the stock keeping preferences of the pharmacist. A product operating in a niche market will have relatively low number of units sold in a brick hence any increase in the prescription preferences of physicians for that product will probably result in earlier orders by the pharmacists to the wholesalers since the stock preference for that product will be at lower levels, which will reflect as a shorter lag.

Due to this lag, our hypothesis needs to accommodate testing of different sales effect scenarios, by which we can identify the actual response lag per each product separately. For this purpose we will include the sales of the product, one, two, and three months after the actual month when the detailing activities are done. Next section will run the statistical analysis to first discover if there's a correlation, and if there is, what the lag for the five products being detailed is. For all our statistical analysis, we will be using the linear regression model.

Hypothesis 1

This analysis will first test the correlation of the sales versus the activity data without going into the details of activities made to different customer segments. The outcome of this first analysis will reveal the lag per product and the correlation factor between activity and sales.

H_0: There is no relationship between any of the previous three months activity counts and market share.

H_A: There is at least one significant relationship between any of the previous three months activity counts and market share.

Hypothesis 1 linear model:

$$MarketShare = \beta_0 + \beta_1 CallCount_{m1} + \beta_2 CallCount_{m2} + \beta_3 CallCount_{m3} + \beta_4 Brick + \epsilon$$

where market share is the percentage share of units in an ATC class of the product of the T-Pharma company, β_0 is the intercept term, β_1 is the slope of the correlation between call count of 1 month before and market share, while m2 and m3 represent the call counts of previous 2nd and 3rd months respectively. Brick is added to the equation as a dummy variable, because each region has a unique sales pattern within itself and must be factored in to the analysis. Since the analysis is done per product, a dummy variable for products was not included in the hypothesis.

Results for Hypothesis 1

The linear model is applied to each five products separately to determine the correlation of the sales activities versus activity. The products are Product-A, Product-B, Product-C, Product-E, and Product-D. Following sections will present the analysis for each product.

Product-A

Table 2. Hypothesis 1 Linear Regression Results for Product-A

Product	IndependentVariable	Estimate	Standard Error	T-Value	P Value
PRODUCT-A	intercept	19.340	3.230	5.987	0.00000000240*
PRODUCT-A	m1	0.014	0.008	1.866	0.06220478037
PRODUCT-A	m2	0.049	0.007	6.597	0.00000000005*
PRODUCT-A	m3	0.053	0.008	6.842	0.00000000001*
				R Squared:	0.43

Product-A has a fluctuating sales during the period of 6 months, some of it due to campaigns run by Pharma-T to promote the sales of the product, thus the analysis cannot healthily predict the relationship between calls and sales and R Squared is fairly low at 0.43. Nevertheless, its statistical analysis provides with some insight on the lag for the sales and activity correlation and identifies the bricks where this correlation was highly observable.

Table 2 shows that the activities made three months prior to the sales, represented by $CallCount_{m3}$, are most effective in predicting the sales of the product with the above model. $CallCount_{m1}$ is found insignificant while $CallCount_{m2}$ is found to be significant but not as high as $CallCount_{m3}$, it may be a good candidate in showing the lag for this product. This shows that the lag of Product-A to responding to call detailing activities in terms of sales is between two or three months ago. The r squared is low due to the aforementioned sales campaigns conducted by T-Pharma during the analyzed period of time.

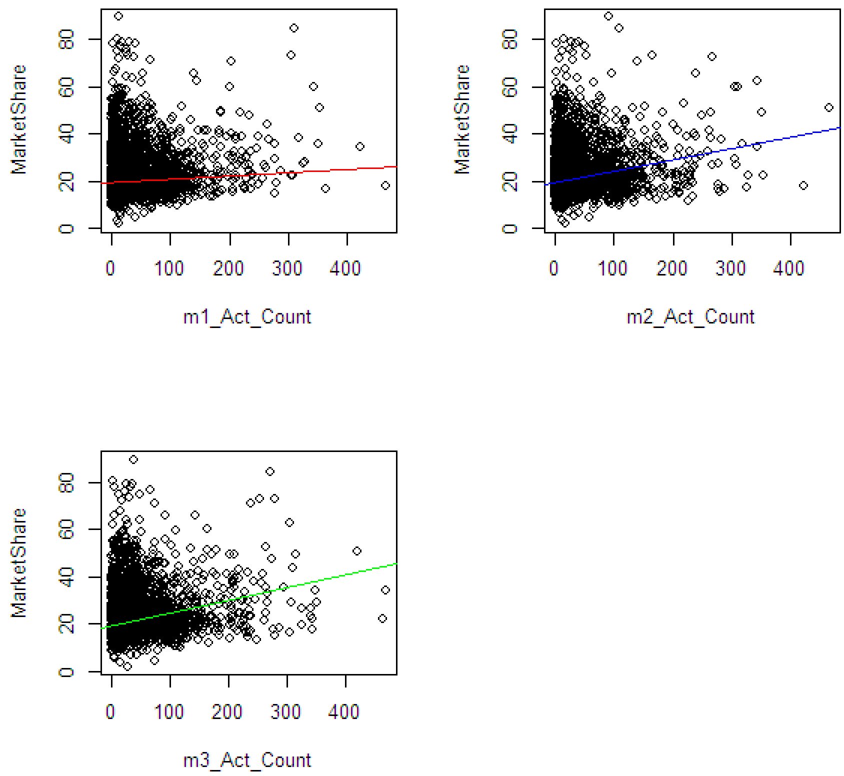

Figure 6. The plot graphs of $CallCount_{m1}$, $CallCount_{m2}$ and $CallCount_{m3}$ against market share

Graphs in Figure 6 show the predicted slopes; β_1, β_2, β_3 respectively against the *y* axis of market share. Since the bricks are factored in to the analysis, looking at the overall data with all bricks included is somewhat misleading to see the relationship which is predicted by the model.

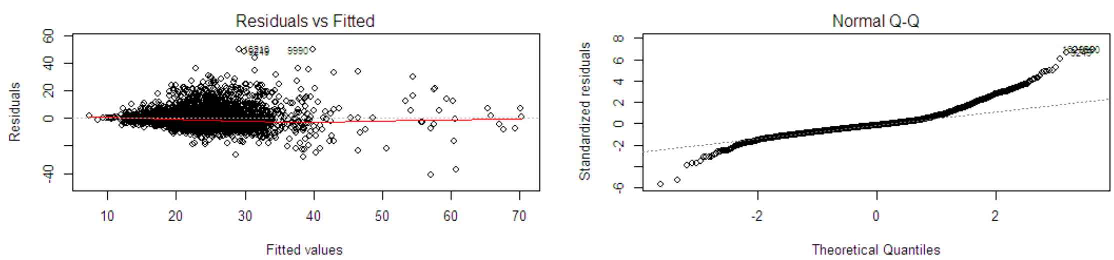

Figure 7. Residuals vs. Fitted and Q-Q Graphs to analyze the validity of the linear model applied

In Figure 7, Residuals vs. Fitted graph, shows no pattern in the fitted values of y hat towards the error terms. Also the Quantile-Quantile (Q-Q) plot shows a normal distribution of the error terms, pointing out conformity with a linear model.

Although Product-A seems to be a good fit for analysis into customer segment level, it lacks a healthy r squared to be trusted on the model. This product will not be used for development of the simulation model.

Product-B

Product-B is one of the leaders in its market and has reached a mature state in terms of its sales and its place in its ATC class. For this main reason, the response of the sales in regards to the detailing activities is fairly low. Even so, the analysis of this product reveals some important insights.

Table 3. Hypothesis 1 Linear Regression Results for Product-B

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-B	intercept	58.678	4.710	12.457	0.0000000001*
PRODUCT-B	m1	0.031	0.012	2.525	0.0116394799*
PRODUCT-B	m2	-0.046	0.013	-3.500	0.0004734369*
PRODUCT-B	m3	-0.058	0.015	-3.937	0.0000848057*
				R Squared:	0.70

Table 3 shows that the activities done two and three months prior to the sales, represented by $CallCount_{m2}$ and $CallCount_{m3}$, are inversely correlated with the market share. This is primarily because there is a coincidental conformity across the sales and the activity trends of the previous months. It is apparent from the analysis that calls done 1 month prior to the sales best defines the response function. The r squared of the model is fairly high at 70% for a set of 2500 rows of data for this product.

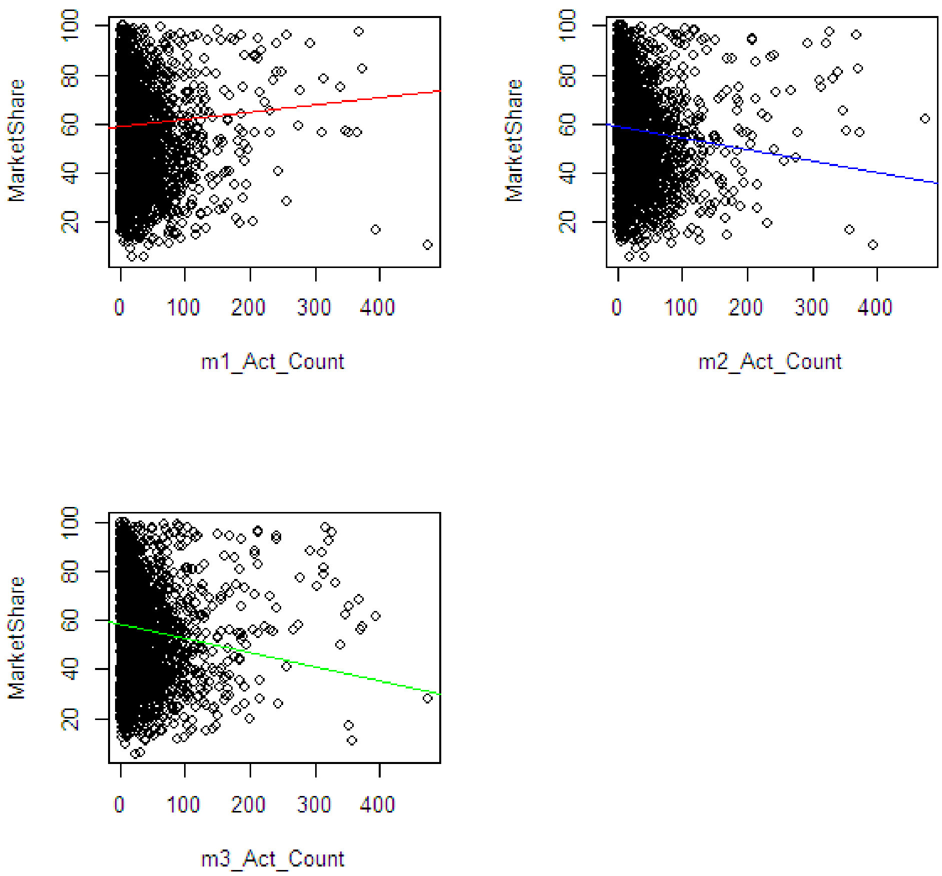

Figure 8. The plot graphs of $CallCount_{m1}$, $CallCount_{m2}$ and $CallCount_{m3}$ against market share

The first graph in Figure 8 shows the probable effect of detailing activities with a positive correlation while the other two graphs display a negative correlation. A downward slope is primarily due to decreasing trends in the market share coincidentally aligning with the decreasing trends of the detailing activities in this six month windows.

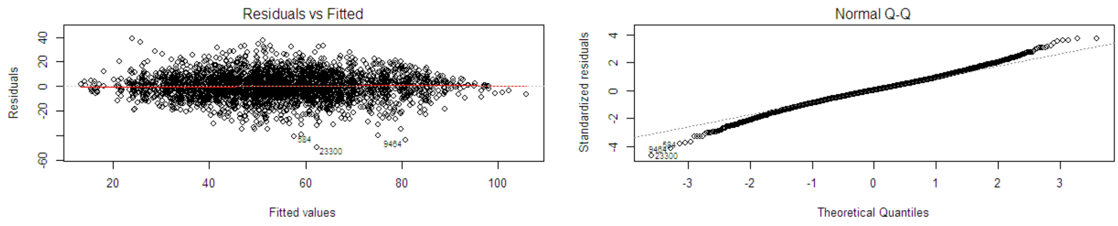

Figure 9. Residuals vs. Fitted and Q-Q Graphs to analyze the validity of the linear model applied

On the other hand, the two graphs in Figure 9show that the fitted model is healthy in terms of normal distribution of error terms and with the lack of correlation between the residuals and the fitted ŷ values.

Product-B is not a good fit for taking as a basis for our simulation building, due to its inverse correlation of previous months and its sales response function, as it is not sensitive because of the maturity of its sales in the market.

Product-C

Product-C is a new product in the market and has a small market share but it has a rising trend. There are many marketing activities supporting the promotion of this product so understanding the relationship and effectiveness of the detailing activities in determining sales will be a challenge. But from another perspective, the product is new and should present better the trend of sales in response to promotions by the field force. The market it operates in is also an emerging market so the market share is highly dependent on the success of its competitors in promotional activities.

Table 4. Hypothesis 1 Linear Regression Results for Product-C

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-C	intercept	6.699	7.510	0.892	0.07251740874
PRODUCT-C	m1	0.051	0.022	2.338	0.01948280249*
PRODUCT-C	m2	0.051	0.021	2.446	0.01453461994*
PRODUCT-C	m3	-0.030	0.025	-1.206	0.22777326914
				R Squared:	0.35

The result of the hypothesis 1 is displayed in Table 4. As the intercept row shows the market share which is the y intercept of the slopes of three months' β slopes, is 6.7, this is the average market share of the product. The model itself is not the best fit with p value of 0.07 shows and the r squared is low at 35%. It is observed that call count in month 1 and month 2 have the same slope with very similar p-values, this is due to a low variance in the call counts across the six month period in which this data is observed. Also the negative correlation of the month 3 variable points out to a flaw in this model.

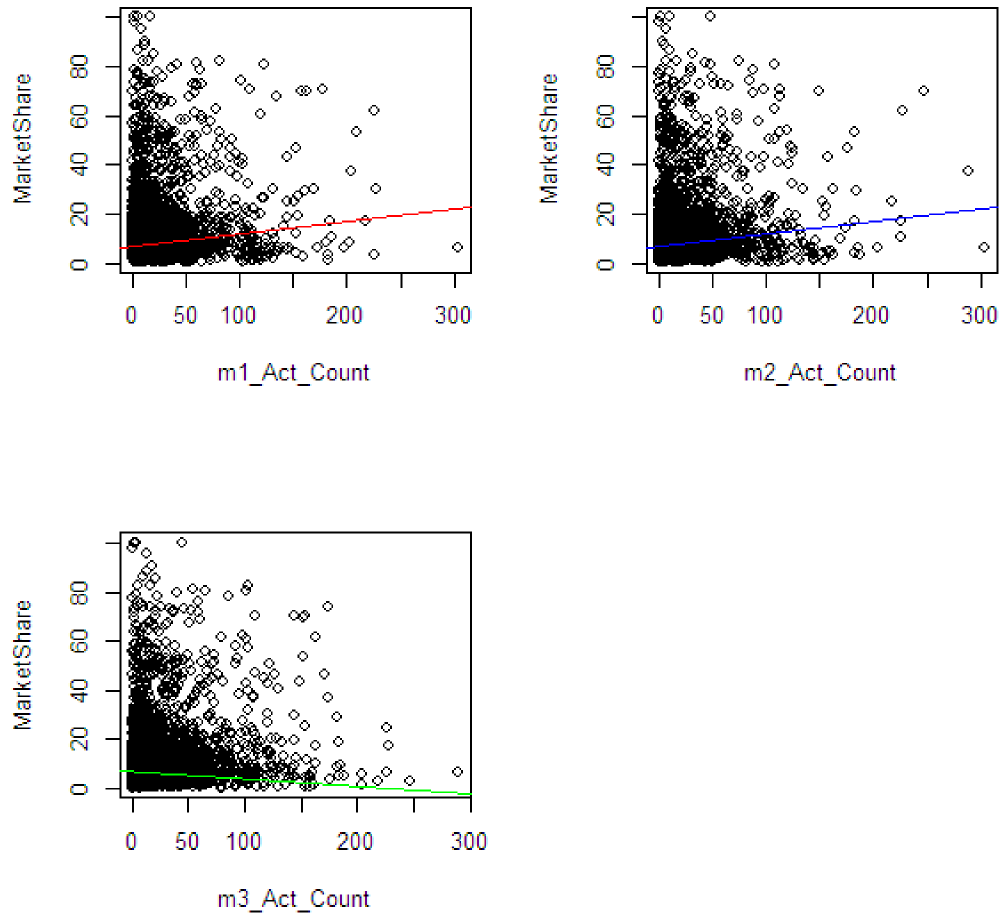

Figure 10. The plot graphs of $CallCount_{m1}$, $CallCount_{m2}$ and $CallCount_{m3}$ against market share

The predicted y values by each slope are displayed in the above graph (Figure 10). The activity counts and market shares are highly concentrated in the 0-100 and 0-40 ranges respectively. This is due to low variance in both dependent and independent variables.

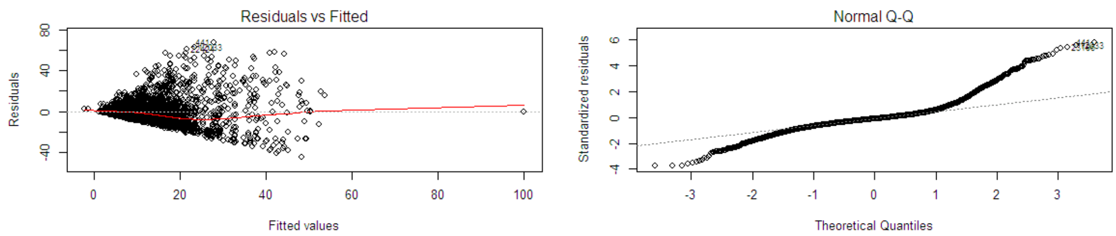

Figure 11. Residuals vs. Fitted and Q-Q Graphs to analyze the validity of the linear model applied

As the residuals vs. fitted graph shows in Figure 11, the data is highly concentrated in a small scope, pointing out to low variance. Also the Q-Q graph does not potentially represent a normal distribution of the error terms and confirms the invalidity of the model.

Although Product-C is a new product and had potential in presenting a healthy linear model for hypothesis 1, it had certain other factors which affected the data. The market that it operates in is an emerging market with new products being launched and these new launches result in market share stealing from other existing products in the market. Also, new products are highly volatile in terms of market share since they will have many marketing channels working for its promotional activities as well as campaigns which promote the sales of the product with narrow timed discounts or sales deals. Product-C is not a good candidate for building our simulation model.

Product-D

Product-D is unique amongst the 5 products analyzed in this section, because T-Pharma has another product competing with her own. Product-D is in its transitioning stage to being a cash cow for the company and the other competitor product from the same company is a new product which joined its portfolio after a merger. Product-D is a chronic disease treatment drug and does not have much seasonality. Also, its customer base is limited to a few specialties and thus the detailing activities are highly important in gaining sales and sustaining its customer base.

Table 5. Hypothesis 1 Linear Regression Results for Product-D

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-D	intercept	24.558	3.147	7.803	0.00000000001*
PRODUCT-D	m1	0.044	0.018	2.493	0.01276158183*
PRODUCT-D	m2	0.002	0.017	0.099	0.92084807041
PRODUCT-D	m3	0.023	0.019	1.215	0.22467606275
				R Squared	0.60

As observed from Table 5, the product averages at 25% market share and the model is found significant with p value of 1e-10 to support the hypothesis. The r squared is 60%, which is fairly high when compared to the lack of direct relationship between activity and sales due many other promotional factors and the competitors' success or failure at their own marketing and sales activities. It is apparent that only the previous month's activities are found to be significantly correlated with market share with a 0.01 p-value. The other independent variables' are not significant and cannot predict the sales value as good as month 1.

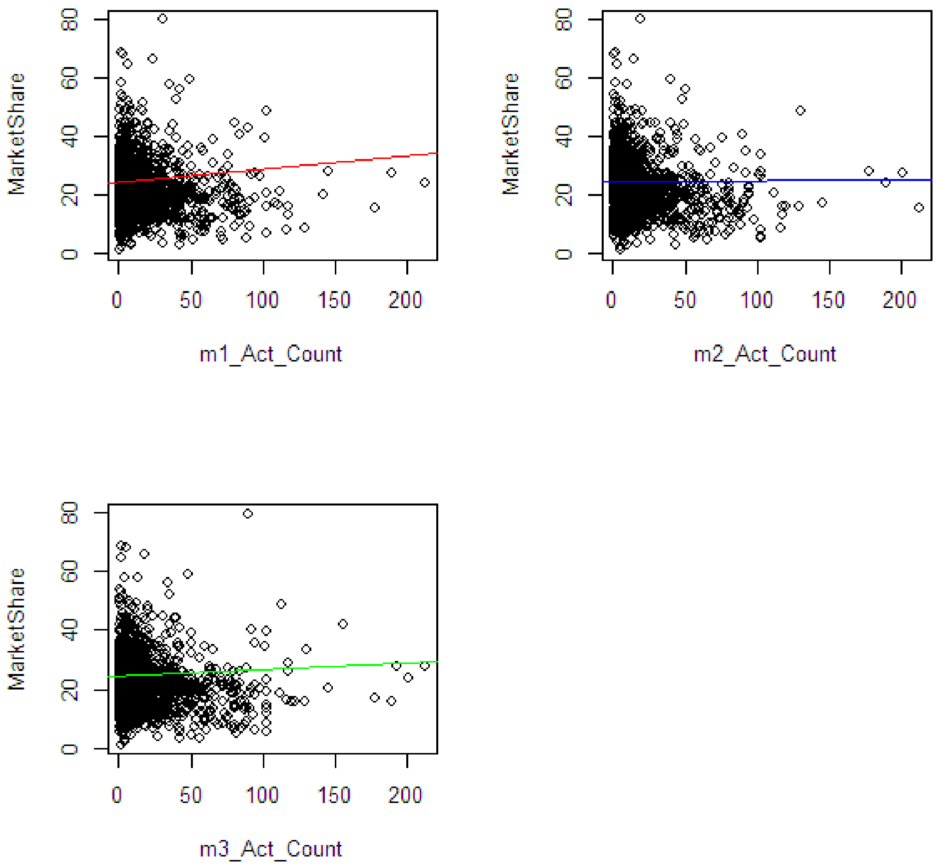

Figure 12. The plot graphs of $CallCount_{m1}$, $CallCount_{m2}$ and $CallCount_{m3}$ against market share

Above in Figure 12, we can observe the predicted linear model for all three different independent variables. Only the first one is found significant and the slope is much higher compared to the other independent variables. This shows high impact of activities, made to customers detailing this product, on sales.

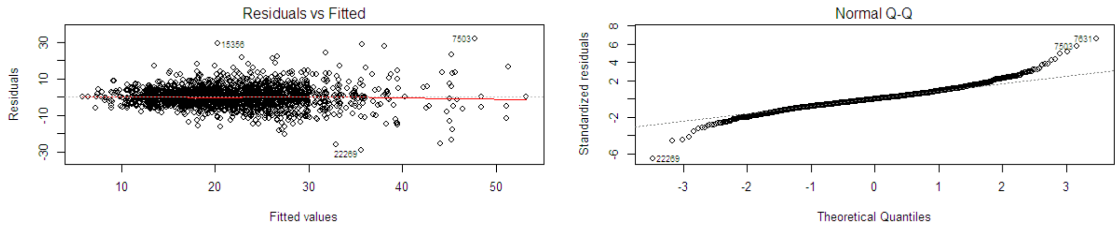

Figure 13. Residuals vs. Fitted and Q-Q Graphs to analyze the validity of the linear model applied

The graphs in Figure 13, show that there's no pattern in the residuals vs. fitted values and that there is a normal distribution across the error terms. These graphs point out to a well fitted and correctly used linear model to test the hypothesis.

Product-D revealed significant results, as it showed a significant correlation between its detailing calls of previous month and its sales. The market share averaging at 24.5% increases over the course of six month aligned with the detailing calls made for that product. The r squared also is also relatively good, but the promotional activities for its rival product from the same company makes it a deficient product to base the simulation on, because the sales force efforts are shifting and are more allocated for the other product as time passes. This creates a decrease in both sales and detailing activities, making them correlated but in reality correlation is not due to the response in sales to activities made, it is because the company is now diverting its energy to the newly launched product.

Product-E

Product-E is a product in an emerging stage but is close to being a stable product in the market. It is a leading product in its market, but the market itself is dynamic and sales activities are found to be significantly effective in this market. It is only prescribed by a few specialist physicians and because the drug has potentially prolonged usage times, new customers who are gained will probably continue using the same product until their illness is cured. Promotional activities by the field force ensure that the messages of the marketing department and the literature of clinical trials reach the customers and play a positive effect towards the preference of this drug in prescriptions.

Table 6. Hypothesis 1 Linear Regression Results for Product-E

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-E	intercept	6.549	0.833	7.863	0.00000000001*
PRODUCT-E	m1	0.005	0.003	1.816	0.06957093725*
PRODUCT-E	m2	0.013	0.003	4.346	0.00001453114*
PRODUCT-E	m3	-0.003	0.003	-0.868	0.38558709434
				R Squared:	0.68

As seen in Table 6, Product-E has an average of 6.5% market share and the intercept is found significant (p=1e-11), along with month one and two slopes significant (p=.06 and p=.00001) in determining the market share with positive slopes. The three months prior activities are found to be negatively correlated but this is highly insignificant due to high values of p at 0.38. The effect of previous month is also not big when looking at the t-value of the analysis. It is apparent that month 2 is the determining factor in sales.

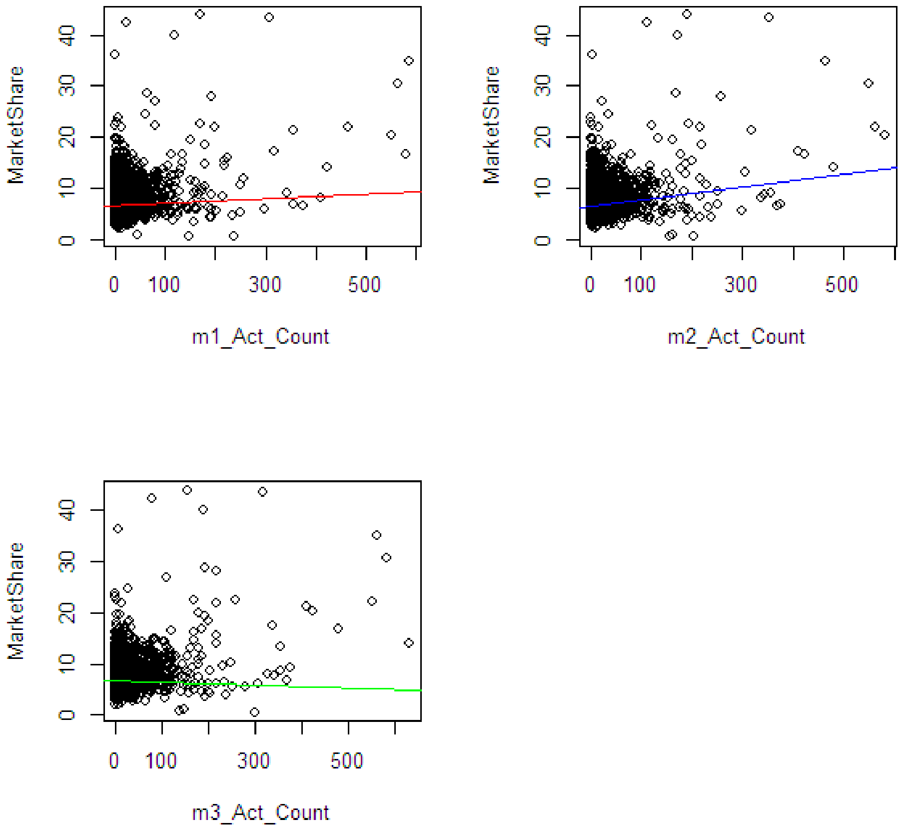

Figure 14. The plot graphs of $CallCount_{m1}$, $CallCount_{m2}$ and $CallCount_{m3}$ against market share

As seen in Figure 14, the significant correlation of calls of month 2 has a high slope which means the responsiveness of sales is high to detailing activities.

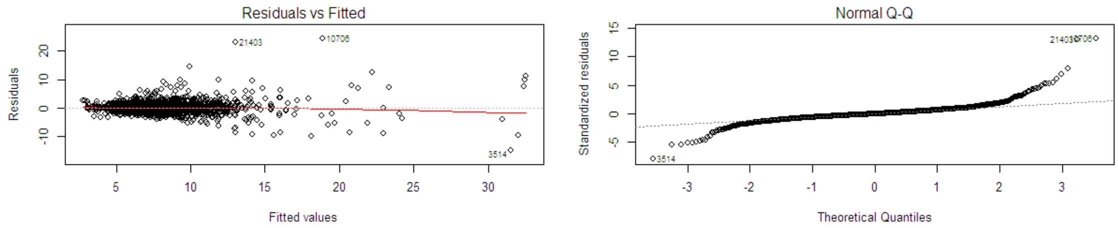

Figure 15. Residuals vs. Fitted and Q-Q Graphs to analyze the validity of the linear model applied

The graphs in Figure 15, show that there's no pattern in the residuals vs. fitted values and that there is a normal distribution across the error terms. These graphs point out to a well fitted and correctly used linear model to test the hypothesis.

Product-E is a good fit for understanding the relations of sales versus detailing activities and analyzing them at the customer segment level. It has a significant and outstanding correlation with the activities made two months prior to the month of the sales measured. It also has a comparatively high r squared at 68%, with normally distributed error terms.

Hypothesis 2

Now that we have analyzed the data for finding the correlation and the lag between the detailing activities and sales of a single product of T-Pharma, we will proceed to second step in understanding the role that each visit to different customer segments play against the success of sales.

Now that the lag is known, a new linear model needs to be developed to understand the effect of visits made to each customer. The activities are divided into 5 segments:

1. *A Segment Customers*: These customers play a highly important role in deriving the sales of a product so the visit frequency and the impact of this visit should be observed higher relative to all other segments.
2. *B Segment Customers*: Since the preference is already set at against the favor of the company, the attainment of trust of these customers is high, but at the same time the effort required to gain their trust is expected also to be high resulting in lower impacting visits to these customers. Although the

importance of visits is high to B customers, the impact of the visits would be expected to be low in principal.

3. *C Segment Customers*: Although the visits to these customers are important to maintain the loyalty of this segment, the impact of the visits would be expected to be lower than the visits made to A customers, but they still would be more impactful than the visits to B customers, again due to tendency of C customers in preferring the product versus the hard to be convinced B customers.
4. *D Segment Customers*: This group has similar traits like B customers but will have lower effect in sales when visited. These customers should be the least visited customers by a sales representative due to the lowest ROI expectancy.
5. *Not Segmented Customers*: Usually this is a group of mixture of all four segmented customers, but since the representatives have more knowledge on A and C customers, usually this group includes presumable more B and D customers. The value of each visit to this group of customers may still expected to be better than the average of what we would expect from B and D customers due to A and C customers in this group. Even though their percentage should be lower in terms of population, because the ROI of visits are expected to much higher than B and D, they should be expected to bring the average higher.

The above assumptions will be tested with Hypothesis 2:

H_0: There is no significant relationship between activities made to different customer segments and market share.

H_A: There is a significant relationship between activities made to different customer segments and market share.

$$MarketShare = \beta_0 + \beta_1 CallCount_A + \beta_2 CallCount_B + \beta_3 CallCount_C + \beta_4 CallCount_D + \beta_5 CallCount_{NoSegment} + \beta_6 Brick + \epsilon$$

The above hypothesis will only be run for the correlation of a single lag which was attained in Hypothesis 1. The analysis will be conducted for each product for the lag of months found to be most correlated with the sales. The purpose of this test will be to find the individual effects of all independent variables of call counts to all five segments described in the above section.

Results for Hypothesis 2

The linear model is applied to each five products separately to determine the correlation of the sales versus activity made to five customer segments. The products are Product-A, Product-B, Product-C, Product-E, and Product-D. Following sections will present the analysis for each product.

Product-A

Table 7. Hypothesis 2 Linear Regression Results for Product-A (Lag of 3 Months)

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-A	intercept	20.356	3.364	6.051	1.63231E-09*
PRODUCT-A	m3_A	-0.090	0.037	-2.450	0.014340325*
PRODUCT-A	m3_B	0.002	0.045	0.054	0.956641816
PRODUCT-A	m3_C	-0.078	0.068	-1.156	0.247942288
PRODUCT-A	m3_D	0.074	0.026	2.845	0.004467392*
PRODUCT-A	m3_NoSeg	-0.039	0.020	-1.927	0.054122205
				R Squared:	0.42

Observing the outcomes of the above t-test results in Table 7, we can conclude that only activities made to A and D customers are significantly correlated with the market share and that the explanatory power of the model is fairly low at 42% value of the r squared. Although Product-A displayed promising results in the first hypothesis, the results of the in depth analysis does not promise a solid model to be used in the simulation.

Product-B

Table 8. Hypothesis 2 Linear Regression Results for Product-B (Lag of 1 Month)

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-B	intercept	58.237	4.799	12.135	6.454426E-33*
PRODUCT-B	m1_A	-0.007	0.107	-0.065	0.947856215
PRODUCT-B	m1_B	-0.150	0.079	-1.905	0.056898507*
PRODUCT-B	m1_C	0.078	0.117	0.667	0.504568199
PRODUCT-B	m1_D	-0.020	0.117	-0.169	0.865636480
PRODUCT-B	m1_NoSeg	-0.062	0.016	-3.878	0.000108082*
				R Squared:	0.70

Product-B was one of the products which had meaningful results in the first hypothesis for lag of one month but had negative correlation with second and third months. In this analysis show in Table 8, we can see that not many significant relatiobship exist with its activities made to customers in different segments and its sales. The model actually finds significant correlation for B and NotSegmented customers, but with a negative correlation. This probably points out to the fact that ROI on these segments, as mentioned before, are expected to be really low, so when the field force actually spends efforts in visiting these customers, they are losing the possibility of raising its market share by visiting the other segmented customers. Product-B's analysis does not reveal any significant results to be taken as a basis for the simulation.

Product-C

Table 9. Hypothesis 2 Linear Regression Results for Product-C (Lag of 2 Months)

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-C	intercept	6.871	7.527	0.913	0.361355773
PRODUCT-C	m2_A	-0.298	0.350	-0.852	0.394539302
PRODUCT-C	m2_B	0.145	0.134	1.089	0.276422230
PRODUCT-C	m2_C	-0.275	0.197	-1.396	0.162777425
PRODUCT-C	m2_D	-0.335	0.246	-1.361	0.173552398
PRODUCT-C	m2_NoSeg	0.014	0.032	0.452	0.651091187
				R Squared:	0.35

Product-C was one of the newly launched products of T-Pharma and as it was discussed in the results of Hypothesis 1, the promotional activities and marketing campaigns are all so intense that it is unable to derive the effect of field force detailing activities on sales. The product also operates in an emerging market so the maturity level of all other products are also fairly low making this market's share values highly volatile. This uncertainty and inconsistency is reflected in the results presented in Table 9. Although the fist model fit for month 1, the deeper analysis into segments of customers reveals insignificant correlation at all levels including the intercept. These results are statistically insignificant and surely cannot be used for the simulation of prescription behaviour and carryover effect.

Product-D

Table 10. Hypothesis 2 Linear Regression Results for Product-D (Lag of 1 Month)

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-D	intercept	25.818	3.162	8.166	6.570267E-16*
PRODUCT-D	m1_A	0.039	0.099	0.396	0.692258724
PRODUCT-D	m1_B	0.066	0.075	0.887	0.375291649
PRODUCT-D	m1_C	-0.078	0.070	-1.105	0.269223889
PRODUCT-D	m1_D	0.039	0.043	0.909	0.363700760
PRODUCT-D	m1_NoSeg	-0.012	0.051	-0.246	0.805466240
				R Squared:	0.60

Product-D's data, as seen in Table 10, reveals all insignificant correlation with all activities made to different segments of customers. This may be due to a poor segmentation of the customers, or due to the fact that there is a competitor product from the same company in the same market and that the company is promoting this new product as opposed to Product-D. This then results in stealing market share of its rivalry, also known as cannibalism in marketing, which results in conversion of existing customer base to the new market. This will then lead to non-correlated variations in the sales and the detailing visits. Product-D is not a good candidate for founding upon the simulation parameters and assumptions.

Product-E

Table 11. Hypothesis 2 Linear Regression Results for Product-E (Lag of 2 Months)

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-E	intercept	6.416	0.831	7.717	1.843031E-14*
PRODUCT-E	m2_A	0.049	0.017	2.897	0.003802393*
PRODUCT-E	m2_B	-0.039	0.017	-2.328	0.019999578*
PRODUCT-E	m2_C	0.054	0.018	2.963	0.003076657*
PRODUCT-E	m2_D	-0.003	0.009	-0.296	0.766945648
PRODUCT-E	m2_NoSeg	0.009	0.005	1.864	0.062410688*
				R Squared:	0.69

Product-E was one of the products which had promising results in the first hypothesis tests. It is a product transitioning to a mature stage but operates in a competitive market. It cures a disease which is treated over longer periods of time and the loyalty of the customer is highly important in expanding market share with new patients prescribed with Product-E. The first analysis had shown that the correlation of activities of two months before with the sales was the most significant. Table 11shows the results of the linear model run against the activities made to different customer segments. The intercept and activities to all segments except D segment is found to statistically significant. The r squared points out that 69% of the data can be explained with this model and t values distribution show contribution of activities made to customers in different segments. The slope of B and D customers are found to be negative. This would mean that the sales representative loses market share when he directs his efforts towards these segments. Although, at first, intuitively this seems abnormal, it is actually an effect observed due to a possible collinearity between the independent variables. It should be noted that the independent variables are actually a distribution of detailing activities to a set of customers in different segments, which means that increasing efforts in one segment would mean to *steal away* effort from the other segments. Since the average impact of visiting all other segments other than B is higher than the effect of the B segment customer itself, the model finds that the market share is lost. We will analyze the meaning of this and how it might be so in the following section.

<u>Testing Mutlicollinearity and Finding Partial Regression Coefficients</u>

Hypothesis 2 revealed plausible results for A and C segmented customers while suggesting negative correlation for B and D segmented customers. We have argued that this is due to the stealing effect of allocating resources to a segment which is not as high as ROI expected from the first two segments. This suggestion can be tested by testing for the multicollinearity of the independent variables. If there is a multicollinearity among the independent variables, this suggestion will be regarded as true. If positive results are achieved from this test, we will continue to finding the partial regression coefficients of the independent variables. The partial regression coefficient may be defined as the net effect of a single independent variable while holding the other variables constant.

To understand the level of multicollinearity among the independent variables we will run an analysis to find the variance inflation factor (VIF). It will show the severity of the multicollinearity by testing the increase in variance of the estimated regression coefficient because of collinearity. If the VIF turns out to be 5 or above, the multicollinearity will be regarded as existing, while numbers above ten will identify severe cases.

Table 12. VIF Values for each Independent Variable for Product-E

Independent Variable	GVIF	Df	GVIF^(1/(2*Df))
m2_A_Act_Count	7.338631	1	2.708991
m2_B_Act_Count	6.331582	1	2.516263
m2_C_Act_Count	10.889259	1	3.299888
m2_D_Act_Count	9.134407	1	3.022318
m2_NoSeg_Act_Count	8.633479	1	2.938278

In Table 12, the values of variance inflation factor for all independent values is above the threshold of 5, with calls on C segmented customers exceeding the limit for severe collinearity.

Now that the assumed multicollinearity is proved with the VIF analysis, we will attain the net effect of each independent variable; namely the call count made to each five segmented customer groups, we will use the partial correlation coefficient methodology. A step wise example is provided below to attain partial regression coefficient for Call Count A:

1. Regress MarketShare with Call Count made to each four segment except for the segment for which we want to achieve the net effect.

   model1 :

$$MarketShare = \beta_0 + \beta_1 CallCount_B + \beta_2 CallCount_C + \beta_3 CallCount_D + \beta_4 CallCount_{NoSegment} + \beta_5 Brick + \epsilon$$

2. Get the residuals of the regression excluding the call count of the segment:

$$\varepsilon_1 = residuals(model1)$$

3. Regress CallCount_A with the same four segments.

   model2 :

$$CallCount_A = \beta_0 + \beta_1 CallCount_B + \beta_2 CallCount_C + \beta_3 CallCount_D$$
$$+ \beta_4 CallCount_{NoSegment} + \beta_5 Brick + \epsilon$$

4. Get the residuals of the regression excluding the call count of the segment:

$$\varepsilon_2 = residuals(model2)$$

5. Now regress ε_1 with ε_2 to get the net effect of CallCount_A in this example:

$$\varepsilon_1 = \beta_0 + \beta_1 \varepsilon_2 + \epsilon$$

With above 5 steps, we will attain with β_1 the net effect of calls made to A customers only to the market share. This method for will be executed for each segment to attain the net effect of visits made to each of the five customer.

Table 13. The Partial Regression Coefficient for all Segments

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value	R Squared
PRODUCT-E	intercept	0.00000	0.05951	2.13240	0.14620	0.43256
PRODUCT-E	A	0.10694	0.00992	10.78138	0.00000001	0.43256
PRODUCT-E	intercept	0.00000	0.05951	1.98212	0.24301	0.07948
PRODUCT-E	B	0.02005	0.01103	4.53845	0.00001	0.07948
PRODUCT-E	intercept	0.00000	0.05951	3.48856	0.09445	0.03957
PRODUCT-E	C	0.03613	0.01131	3.19587	0.00141	0.03957
PRODUCT-E	intercept	0.00000	0.05951	0.95332	0.64212	0.06742
PRODUCT-E	D	0.02515	0.00602	4.17742	0.00003	0.06742
PRODUCT-E	intercept	0.00000	0.05951	1.53221	0.42366	0.13103
PRODUCT-E	NoSeg	0.01948	0.00333	5.84247	0.000000001	0.13103

The regression between the residuals of each partial regression in Table 13 has revealed not many significant results but gives an idea on the relationship of the effect of each visit to different segments on the output of market share. It is apparent from the Table 13 that A has the highest influence while non-segmented customers have the lowest.

Hypothesis 3

Because there is multicollinearity amongst the independent variables and because prescription characteristics of some classes are similar, we will reduce the number of independent variables and group segments of customers. A and C customers are the customers where market share of the product of T-Pharma is high and for B and D customers it is low. For not segmented customers, there is no survey data and we can assume that these customers are not of importance for T-Pharma, therefore they can be either B or D. So we will group A and C together, while grouping B, D and No-Segment customers together. Our new hypothesis for understanding the effect of these customer segments will be, as Hypothesis 3:

H_0: There is no significant relationship between activities made to AC and B-D-NoSegment customer segments and market share

H_A: There is a significant relationship between activities made to AC and B-D-NoSegment customer segments and market share.

Hypothesis 3 model:

$$MarketShare = \beta_0 + \beta_1 CallCount_{AC} + \beta_2 CallCount_{BDNoSegment} + \beta_6 Brick + \epsilon$$

In this way, we will have grouped the calls made to A and C customers and use them as a single independent variable, and same applied for the other three segments. Brick stays as a dummy variable in our linear regression model.

Below in Table 14, the results for hypothesis 3 are summarized:

Table 14. Hypothesis 3 Results for Product-E.

Product	Independent Variable	Estimate	Standard Error	T-Value	P Value
PRODUCT-E	intercept	6.445	0.83	6.051	1.431E-14*
PRODUCT-E	m2_AC	0.0454	0.011	3.85	0.011532*
PRODUCT-E	m2_BDNoSegment	0.0218	0.003	1.96	0.05279*
				R Squared:	0.76

The regression is found to be statistically significant with a p value of 1.43E-14 for the intercept term. It is apparent from the results that calls made to A and C customers combined have 2.08 times the effect of calls made to B, D and No-Segment customers. This will be the primary multiplier we will use in building the simulation model, which will be described in next chapter.

CHAPTER 4

SIMULATION MODEL

This chapter will define the basis of the data and methodology used in building the agent based model, describe the details of the simulation and conclude with results and outputs from the simulation.

It has been discussed previously that the carryover effect of the detailing activities made to physicians has a lasting effect (Taneja, 2008). We will set out to discover the factors included in this carryover effect in two criteria:

1. Memory effect: this is the effect of a single visit on the prescription preference of the physician, measured in favour of T-Pharma co.
2. Decay factor: this represents the carryover or in other words the lasting effect of a visit. Decay factor represents the rate of decrease in the prescription preference of the physician during the time where no other promotional activity or detailing visit is made.

Our model will set out to discover these two parameters for different bricks. We will be analyzing bricks separately, since we have included them as a factor analysis in our statistical models. Each brick has a different starting market share in the analyzed six months, but has the same slope for visits made to each customer segment, representing the effect of each visit on the output, which is market share. We have also found the relationship of the effect of A and C customers versus the B, D and No-Segment customers. The multiplier, which is 2.08, will be used in the model as the relationship of memory effects of the customers belonging to these segments. Now we will try to develop a model which works on a visit-to-visit basis, calculating each day gain or loss in the market share attained by the visits made in that day to different customer segments and by closely measuring the carryover effect and the memory effect of each visit. Before we go on to detailing out the model we will need to explore the basis of our model and the parameters it uses and the output it produces.

This section will summarize the model used in the simulation and the foundation it sits on while building the simulation. It will explore the reasons and decision points in arriving to this model while giving further information about the data it will use while building them.

Survey Data for Customer Classification

As explained in Chapter 2, a survey is conducted for classification of customers into segments of A, B, C, D. This classification is based on a simple survey conducted by the T-Pharma company to its sales force using its sales force automation tool. The survey, although measures many different aspects, asks two main questions for determining the segments of each physician in a sales representative's territory. Remember that each sales representative works in a single team, which depicts which products he/she details to his/her customers. Thus, there's only one sales representative to answer survey questions related to a single product. For determining the segmentation of the product, there are two questions asked in this survey:

1. How many units of drugs the customer prescribes in one month?

2. How many units of drugs of the company T-Pharma does the customer prescribe in one month?

While the first question determines the sales output of the customer for the symptom or ATC class of the product in questions, the other determines the amount of units of products prescribed by the customer for T-Pharma Company. With dividing the answer to the second question by the answer to the first question, we can arrive at the market share of the product for each individual physician.

After all sales representatives fill out the survey, for each product they detail, and for most of the physicians in their territory, they are analyzed by the sales force effectiveness department in the organization to determine the cut off points for each segment A, B, C, D. The cut off points for both questions is 30%. The customers who are renounced as non-segmented customers in this study are the ones where no survey information exists for that customer for the product in question, this is either due to negligence of the sales representative or simply because he/she does not have any information on these questions for these physicians. The two questions used in this survey will be used as a basis for the simulation model in determining the prescription behaviour of a physician depending on its segmentation.

Calling on Physicians, Prescription Behaviour and the Carryover Effect

In Chapter 2, we have detailed out how pharmaceutical companies position their sales organizations in promotional activities, but we have not covered the day-to-day calls and how these calls may affect the prescription preference of physicians in detail. Since our simulation model will be based on days and will simulate a real life environment with each physician individually represented by an agent in the simulation model, it will be important to set forth the environment in which the sales representatives operate.

A sales representative first plans their route for a week or a month, usually with the aid of CRM or SFA systems (Agnetis et al., 2010; Baier et al., 2012). Then on a daily basis, the sales representative executes these calls and travels every day to visit the customers he/she has targeted in advance. When a visit is made to a physician, usually the sales representative answers questions about the drug's latest clinical trial results, helps the physician find out resource in the medical literature, helps the physician get in contact with another physician or medical agent for answering their question, or simply discusses a latest update or finding about the product (Gönül et al., 2001; Puschmann & Alt, 2001). These calls may have different effects due to its content, its success in delivering what was asked by the physician or due to effectiveness of the sales representatives but it has been found that the effect of a call usually averages out over time and only varies by each customer segment (Gönül et al., 2001).

When the sales representative finishes calls in one day, he then would continue with the next day with new customers in his target list. It is a usual case for the sales representative to visit same customers more than once in one month due to the importance of the customer and due to the fact that some calls need a follow up call to answer physician’s needs. It is again found from the analysis of the call data for T-Pharma that customers in the same segment also differ in the visit frequency they get. This is probably due to favourite customers of sales representatives and their tendency to keep closer relationships with some of the customers. Even though these customers with different visit frequency may be in the same segment, the sales representative may choose to visit some of these customers more than others. It has been also observed in some rare cases that some of the customers in important segments like A customers, which is the segment where physicians prescribe over average and the market share of T-Pharma is highest, are never visited for a whole month, sometimes for several months. This may be due to three reasons; first reason may be that the sales representative is visiting them and not reporting that visit in the SFA system, second reason may be that the sales representative may have provided wrong information in the survey about the physician and the customer is not actually an A customer, or a third reason may be that the physician already chooses to prescribe the drug of T-Pharma even without any detailing activities made to them. Also, the visit pattern highly depends on the proximity of customers in an area rather than their segment. The sales representative may drive to a location one day and visit all customers in that location, regardless of their segment. Also the availability of the customers is also a restriction and even though the customer may be important, the rep may not be able to coincide with their available time and not be able to visit them. These sporadic and mixed visit behaviours of the sales representatives are hard to formulize and set up in a simulation environment. We will need to simplify this in our model. We will be randomly selecting each customer to visit and run repetitions to get an average output of different visit patterns from the simulation.

After each visit, it has been observed that the prescription preferences of physicians generally turn in advantage of the product detailed (Al-Hamdi et al., 2012; Wood, Gumbhir, Anderson, & Anderson, 1992). Of course this would mostly be observed when the call achieves its purpose and is successful, but for the purpose of simplification, we will assume that all visits are equally successful and yield the same effect. After the sales representative completes a call, he/she may not re-visit the same doctor for a while, depending on the segment of the customer and the importance the sales representative pays to the customer and the route of the sales representative. It has been observed in T-Pharma that Asegment customers are usually visited two or three times in a month while the lowest segment D customers are visited as seldom as once every quarter.

During the interval when the doctor is not visited, it has been observed that the effect of the last visit continues but decreases steadily(Gönül et al., 2001). This slowly decreasing but long lasting effect is called the carryover effect. Conceptually it represents two values; the effect of the visit in affecting the customer's behaviour in favour of the company and the decrease over time but the long lasting effect of that preference change. As detailed in Chapter 2, there have been few studies which tried to address and model the effect and its decay over time, using different methodologies. These include building neural networks to understand these functions, using non-linear statistical functions or some have preferred to use heuristic approaches in determining these factors (Agnetis et al., 2010; Al-Hamdi et al., 2012; Eccles, Grimshaw, Walker, Johnston, & Pitts, 2005; Taneja, 2008; Wright & Lundstrom, 2004; Yi et al., 2003; Yi, 2008).

In this thesis we will use the method of agent based modeling in achieving and simulating these parameters. We will focus on a single product; Product-E, which has proved to be statistically significant in Hypothesis 3, while constructing and testing our simulation model. We will be using NetLogo 5.0.5 in developing the simulation.

Development of the Simulation and its Workflow

The simulation we will develop will focus on a single sales representative and a single brick. We will use data for one product, Product-E, as discussed before, for a six month period. The model will take in values to understand how many doctors are present in a brick and what their prescribing behaviours are. The model will run the actual visits made by a sales representative in the 6 months - this is the same data used in the statistical analysis in Chapter 3 - and will try to find the optimum values for the memory effect and decay factor of different segments of customers which would reflect as closely as possible the values attained in real life as market share. A lag of two months discovered in the statistical analysis is factored in when matching the output of market share, so the model will not generate the market share in two months, rather the market share at the end of the month will represent the market share attained two months later than those visit patterns were made by the sales representative. To simplify the model, we will only create two memory effect and decay factors:

1. Memory effect and decay factor for A-C: the memory and decay factor of A and C will be assumed to be the same. This is due to the fact that loyalty of these customers to T-Pharma is considered to be equal although the total number of prescribed drugs is different. The loyalty represents the likelihood of the customer to prefer the company drug over the competitor drugs and this would indicate that the memory effect - which is the impact of a single visit - and the decay effect - which is the carry over effect of a visit - should be similar in both segmented customer groups.

2. Memory and decay factor for B-D-NoSeg: The assumption made in the first point applies to this second set of output parameters measured. Because the loyalty of B and D segmented customers are low, it is expected that the impact of a single visit made to these customers should be similar and lower than the A and C customers. It would also be plausible to assume that same rules would apply for the carryover effect of the visit to be similar for B and D customers and that it should be different than of A and C customers. The memory effect of B-D-NoSeg customers will be defined as memory-A-C / 2.08. This is the number attained from the Hypothesis 3 results in Chapter 3.

Now let us construct the model and describe how it is shaped. Firstly, let us consider the agents that will be used in the NetLogo simulation:

- Turtles: there will only be one turtle from the beginning to the end of the simulation, representing the sales representative in the region. The turtle will simulate visiting doctors (patches) and affect the memory of the visited doctor. The sales representative will choose the next doctor to visit based on the last visit time of the doctor and how many visits left in that month.
- Patches: the patches will represent each individual doctor in a brick. The A, B, C, D and NoSegment customers will be represented by yellow, red, green, blue and grey respectively. The other patches will be painted to black and they will have no effect to the simulation; they will only remain as non-visited, non-prescribing patches. Each patch will have a memory parameter. This parameter is the determining factor of how many units of drugs this

individual doctor will prescribe when a day passes and it will decay over time with the designated decay factor of the customer's segment.

Secondly, we will need a list of parameters as the basis for our simulation. These parameters will be as follows:

- Number of A, B, C, D, and NoSegment customers: these are the number of customers in the simulated brick. NetLogo is instructed to create as many patches as designated in this input. Each patch will represent the segment of the physician with the colours assigned to them.
- Rx Total for A, B, C, D, and NoSegment customers: this is the number achieved from the survey conducted on sales representatives to understand the prescriptive behaviours of the physicians. It represents how many units of drugs each individual physician from the same segment prescribes in one month. These numbers are aggregated for each segment at the brick level. For example, if there are five A Customers in a single brick to be simulated, the average of these five A segment doctors will be entered as Rx Total A parameter. If these five doctors prescribe 200, 300, 250, 220 and 240, the value for Rx Total A will be 242 units of drugs per month. This value will be calculated for each customer segment in the brick and provided as a parameter to the simulation.
- Rx Product for A, B, C, D, and NoSegment customers: this is the number achieved from the survey conducted on sales representatives to understand the prescriptive behaviours of the physicians, this time towards their preference of the product of T-Pharma. It represents how many units of the product of T-Pharma each individual physician from the same segment prescribes in one month. These numbers are aggregated for each segment at

the brick level. For example, if there are three C Customers in a single brick to be simulated, the average of these three doctors will be entered as Rx Product C parameter. If these three C segment doctors prescribe 22, 41 and 17, the value for Rx Product C will be 26.7 units of drugs per month. This value will be calculated for each segment in the brick and entered into the simulation.

- Visit per month A, B, C, D, and NoSegment customers: this is an array of length six, indicating how many calls were made by the sales representative in the six months to each individual physician in each segment. It might be that there no visits for any of one of the segments of customers, thus the sales representative (the turtle in NetLogo) will not visit any customer while simulating that month.

Because there is no survey information for the Non-Segmented customers, we will use the average of B, C and D segmented customers in the region for calculating the Rx Total and Rx Product parameters for the No-Segment customers. This is due to the fact that if non-segmented customers were A, they would surely be known by the sales representative and would be included in the survey. Because the sales representative did not provide any information about these customers, they should either be not much of importance for T-PharmaCompany or they should be prescribing lower than average amounts of drugs and may be ignored. These assumptions lead to us to argue that they should either be counted as B, C or D segmented customers, if we were to have any means of gathering information about their prescriptive behaviours.

Algorithm of the Simulation

The basis of the simulation is its parameters it takes, as explained in the previous section. The main aim of the simulation is to be able to find the optimum value for the three parameters which would reflect the market share values as close as possible as they are in real life data. The three parameters are:

1. Memory effect alpha: this is the memory effect which is the percentage multiplier that a visited doctor would achieve when visited. The memory effect of B, D and NoSegment customers will be the same as memory effect alpha, whereas the memory effect of A and C will 2.08 * alpha. The memory parameter of the patch will be multiplied by its memory effect value when visited by the sales representative, as determined by its segment.
2. Decay effect A-C: this is the percentage multiplier which would be applied to all doctors of segment A and C at the end of each day. All patches of A and C will have their individual memory attribute multiplied by this decay factor to simulate the forgetting effect of the physician.
3. Decay effect B-D-NoSeg: this is the percentage multiplier which would be applied to all doctors of segment B, D and Not-Segmented at the end of each day. All patches of B, D and Not-Segmented will have their individual memory attribute multiplied by this decay factor to simulate the forgetting effect of the physician.

The simulation will be set to run a repetition of 20 runs, trying out different combinations of these parameters and at the end of allrunswill record the average market share of the six months as output. It will run as many runs as the possible combinations of the parameters in the provided range. We will be searching in the range of 0.0 – 0.5 with 0.01 increments for the memory alpha factor and searching in the range of 0.7 – 1 with 0.02 increments for the decay factor. This would demand for 13056 runs, each with 20 repetitions.

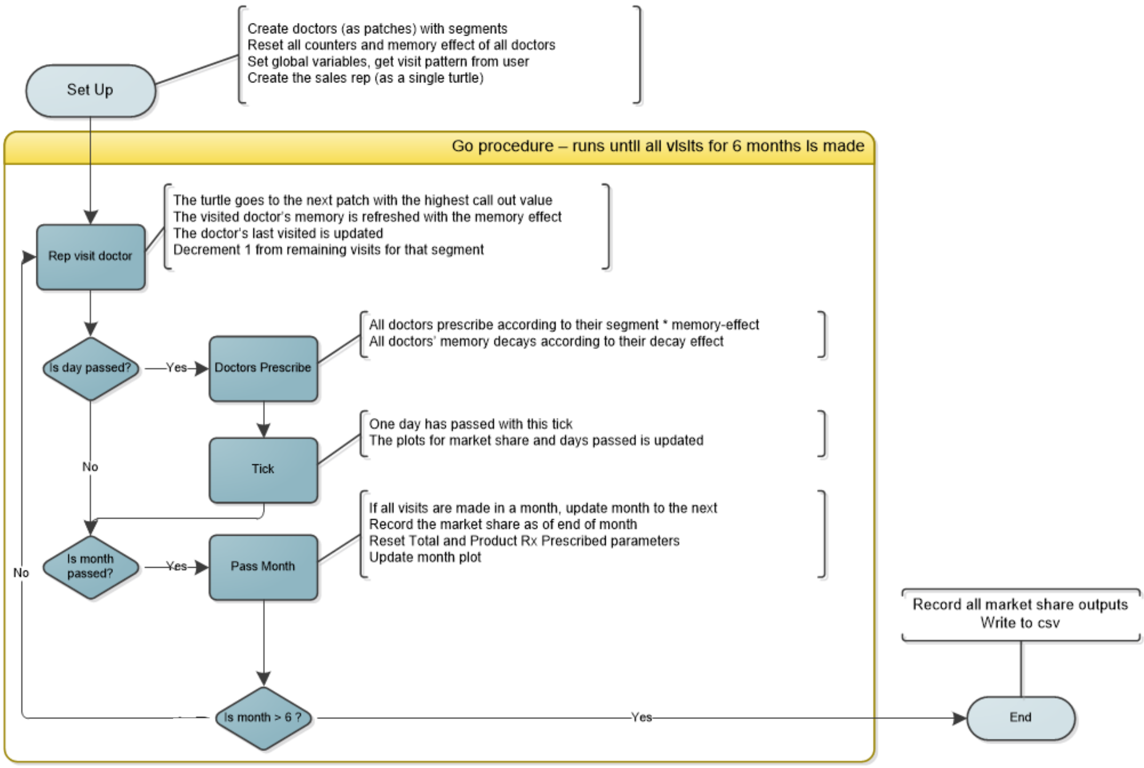

Figure 16. The workflow of the algorithm for the simulation.

The above figure summarizes the overall algorithm workflow of the simulation. Below, a more detailed step wise explanation is presented:

1. Setup: this is where all provided parameters are stored at global variables and all memory of all doctors is set to one. Details of this step are:
   a. Create patches as doctors designated in customer counts provided for each segment. Assign each individual doctor with its defined colour and set initial memory to 1.
   b. Set Rx habits of doctors: each individual patch (doctor) is assigned an attribute based on its segment named Rx Product and Rx Total. These attributes are inherited from the segment information provided via the

user interface. These attributes will determine how many units of drugs this doctor will prescribe and at the start, how many of them will be of the product belonging to T-Pharma company. The memory attribute, which will decay over time when not visited and which will be multiplied by the memory effect multiplier of the segment of the customer when visited, will be used for determining how many units of the T-Pharma product the doctor will prescribe.

c. All global variables are set to their initial values, as entered from the user interface. A month is represented by 20 working days. The prescriptive behaviours of doctors, the memory effect and decay parameters, and the visit pattern for all six months.
d. A turtle agent is created to represent the sales representative.
e. All counters are reset to zero. All plots and graphs are initialized to their starting states.

2. Go procedure: this function is a recurring function, which will iterate until a stop stage is achieved. In this simulation, the stop stage is achieved when all visits are made for all six months. The turtle agent (sales representative) will land on each patch (doctors) and affect the memory of those visited doctors. The doctors will prescribe at the end of each day. The sales representativewill continue to make visits until a month's visit pattern is completely executed. Then at the end of each month, the procedure will store the market share in an array. This output will reflect the effect of the four parameter settings, the memory and decay factors. If these parameters are unchanged, the simulation will always produce the same six month market share. The approach will be to try different combinations of these parameters and find the optimum

settings which would produce the market share as close to as the real life data. The details of the go procedure are as follows:

a. Sales representative visits a doctor: the turtle will choose to move to the next patch with the below function, where last-visited is the ticks value of the last visit, -1 if not visited at all and the visit-at-freq is an attribute to alter the priority that each segment would get in taking visits. Currently the priority is set to be equal for all customers. Visit-left represents how many visits are left to be made in the current month for the segment of that customer.

   move-to min-one-of (doctors with [visit-left > 0])
   [visit-at-freq - (ticks - last-visited)]

b. The doctor executes tasks when visited: the memory attribute of the patch visited will be multiplied by the memory effect parameter of the segment of the customer. Also last-visited is set to the tick value of the simulation. The ticks represent how many days have passed from the start of the simulation. Remember that each 20 days represent one month.

c. Update visits left: decrement 1 from the visits-left parameter of that segment. Once this parameter is 0 for that month, the sales representative does not visit any other customers belonging to that segment until next month starts.

d. Check if a day has passed: each day is represented by a count of visits. This depends on how many visits there are in that month. The condition to have a day passed is:

   *if (visits-made >= ((item month visit-total) / days-in-month) **
   ((remainder (ticks) days-in-month) + 1))

Here, the days-in-month is 20 and the visit total of the month is how many total visits need to be simulated in the current month. If there 60 visits in a month, a day passed after each 3 visits are made. When the day is incremented, a series of important procedures are executed:

i. All doctors prescribe: all doctors prescribe Rx Total amount of drugs, while they prescribe *Rx Product * memory* where memory is the attribute affected by a visit and which decays over time by the decay factor. The global variables Total Drugs Prescribed and the Total Product Prescribed are updated by summing all individual parameters of all doctors with the effect of the memory multiplier.

ii. Every doctor's memory is updated with the decay parameter of their segment; this is the memory decay effect of the carryover.

iii. Market share is calculated from the global variables with the formula:

   (Total Product Prescribed / Total Drugs Prescribed) %

iv. Ticks are incremented by 1, representing a day being passed.

e. Check if a month has passed: with this function we see if all visits for all customer segments are made in the current month. If yes, the below procedures are executed:

   i. Calculate the current market share with the same formula as:

   (Total Product Prescribed / Total Drugs Prescribed) %

   And record in the market share array.

ii. Reset Total Product Prescribed and Total Drugs Prescribed to 0, as we are starting a new month.

iii. Increment the month parameter by 1.

f. Stop if Month > 6: In this step we will check if 6 months have passed and will stop the simulation if it has and present the market-share array as the output of the simulation.

Repetitions and Evaluation Methodology

The simulation runs only for a single brick for a period of six months. In our statistically analyzed data we have 723 in total, but not all of them were found as statistically significant as others. Thus, we will be selecting the bricks which best represent the statistical model and have enough diversity of visits that would help capture the correct memory and decay effect for the five different segments of customers.

The bricks are each represented with a number. In our analysis, the bricks we will analyze are numbered as: 84, 112, 139, 383, 449, and 627. In the next section of outputs, some of these bricks' data will be presented and analyzed. The rest of the results can be found in Appendix A.

The approach for selecting the optimum three parameters is by running a repetition of the simulation with combination of different settings of those parameters within a range and recording each run's output of six month market share in a file. Then we regress that with the real life data and find the output with the least squares of error terms. For example, consider below the output of a run for the brick 112:

Table 15. The Sample Output for Running the Simulation Model with Different Parameters for Brick 112

Run number	Alpha	Memory A-C	Memory B-D-NoSeg	Decay B-D-NoSeg	Decay A-C	[Final ms output]
1	0.01	1.0208	1.01	0.99	0.991	[9.65 5.7 3.84 2.88 2.3 1.92]
2	0. 01	1.0208	1.01	0.99	0.992	[9.75 5.93 4.04 3.03 2.43 2.02]
3	0. 01	1.0208	1.01	0.99	0.993	[9.88 6.22 4.33 3.26 2.61 2.17]
4	0. 01	1.0208	1.01	0.99	0.994	[10.27 7 5.25 4.06 3.26 2.72]
...n						

The output seen in Table 15, is an output produced by running a the simulation with combinations of all three variables which is 13056 runs (depicted with n in the table), each with have 20 repetitions. After all 13056 runs are completed, the output, as captured in the table in the last column named final ms output, is stored in a file. These runs are assigned with *run numbers* and each run number represents a different combination of parameters used in that run. From the table we can see that memory-A-C was 1.0208 because alpha was 0.01 – which is the same value for memory-B-D, for the first run. The least squares method is applied by finding the error term for each month (m1 – m6) and getting the sum of all errors, from which we select the run with the least sum of errors. An example is presented in the below table from the same run of repetitions for the brick 112:

Table 16. The Least Sum of Squares of Errors for Outputs of the Simulation Model for Brick 112

Run number	m1	m2	m3	m4	m5	m6	Real m1	Real m2	Real m3	Real m4	Real m5	Real m6	err_1	err_2	err_3	err_4	err_5	err_6	Sum of errors squared
248	13.0	12.1	11.8	12.0	13.2	11.8	12.7	13.3	11.5	11.7	14.9	11.0	0.1	1.4	0.1	0.1	2.8	0.7	5.1
256	13.3	12.4	12.2	12.0	13.4	12.5	12.7	13.3	11.5	11.7	14.9	11.0	0.3	0.7	0.4	0.1	2.3	2.2	6.0
76	13.2	13.0	12.7	12.4	12.4	12.2	12.7	13.3	11.5	11.7	14.9	11.0	0.3	0.1	1.3	0.4	6.3	1.4	9.8
1063	13.1	12.3	12.4	12.5	14.6	13.9	12.7	13.3	11.5	11.7	14.9	11.0	0.2	0.9	0.7	0.6	0.1	8.2	10.7
883	13.0	12.6	12.1	11.8	11.8	11.6	12.7	13.3	11.5	11.7	14.9	11.0	0.1	0.5	0.4	0.0	9.9	0.3	11.2
1684	12.9	12.5	13.3	13.9	13.2	11.2	12.7	13.3	11.5	11.7	14.9	11.0	0.0	0.5	3.0	4.8	2.9	0.0	11.3

In Table 16, we can see the best six runs whose outputs have the least sum of error squares compared to the real life market share values. The best fit of run identified with the number 248. If we go and check this run from all the run outputs of the repetitions made, we will find the below values in Table 17 for our three parameters:

Table 17. The Optimum Output for the Simulation Model of Brick 112

Run number	248
Alpha	0.021
Memory A-C (derived from alpha)	1.044
Memory B-D-NoSeg (derived from alpha)	1.021
Decay B-D-NoSeg	0.99
Decay A-C	0.995
[Final ms output]	[12.98 12.08 11.82 12.02 13.23 11.83]
Sum of errors squared	5.1412

This same approach will be applied for all bricks to arrive at the optimum value settings for the three parameters in question. The ranges of these parameters are always the same, covering a vast amount of combinations. All of these settings will also be captured in the preceding outputs and findings section.

Outputs, Findings and Limitations Of The Model

In this section, we will present the results of simulations, ran with repetitions for each combination for different bricks and the analysis of the outcomes. These outputs will then be discussed in contrast to each other at the end of the findings section. Finally we will conclude by discussing the limitations of the simulation model and further enhancement possibilities.

Outputs and Findings for the Selected Bricks

The model has run for several repetitions for the below bricks with the detailed parameter ranges and combinations, achieving several outputs. These are then compared to real life data in the approached that was explained previously and the optimum settings for alpha and the decay parameters are achieved. Below we will consider outputs for some of the bricks analyzed, the rest of the outputs can be found in Appendix A of this thesis.

Brick 139

We have searched in the range of 0.0 – 0.5 with 0.01 increments for the memory alpha factor and in the range of 0.7 – 1 with 0.02 increments for the decay factors. Each run has been done with 20 repetitions and the random market share of all 20 runs is taken for the 6 months. The best suited parameter combination is as shown below in Table 18.

Table 18. Best Fit Parameters for Brick 139

Run number	3565
Alpha	9.3897
Memory A-C	0.13
Memory B-D-NoSeg	1.2704
Decay B-D-NoSeg	1.13
Decay A-C	0.94
[Final ms output]	0.98
Sum of errors squared	[10.97 11.14 11.35 10.56 9.65 9.02]

The results reveal that AC customers have a larger decay factor than the B-D-NoSeg customers, and of course due to alpha constraints the memory effect is 2.08 times that of B-D-NoSeg customers. This means that the effect of a detailing activity may be lower for the latter customer segment but the effect lasts longer, revealing a higher carryover factor for this group. This may be counter-intuitive at first but when thought of how rare the customers in segment B, D and No-Segment are visited compared to A and C customers in general, the carryover may be expected to be higher because of the sporadic visit frequency of the sales representative and the sensitivity of the physician to each visit.

The sum of errors squared is 9.38, which means that a monthly average of 1.56 errors squared exists. If we take the square root of 1.56, we achieve 1.25, which is the scalar monthly deviation from the actual market share. The average market share in this Brick is 10.32, so the percentage level variation is 12%, which should reveal a close enough fit to the real life values. When we graph the best fit with the real life values, we achieve the output in Figure 17.

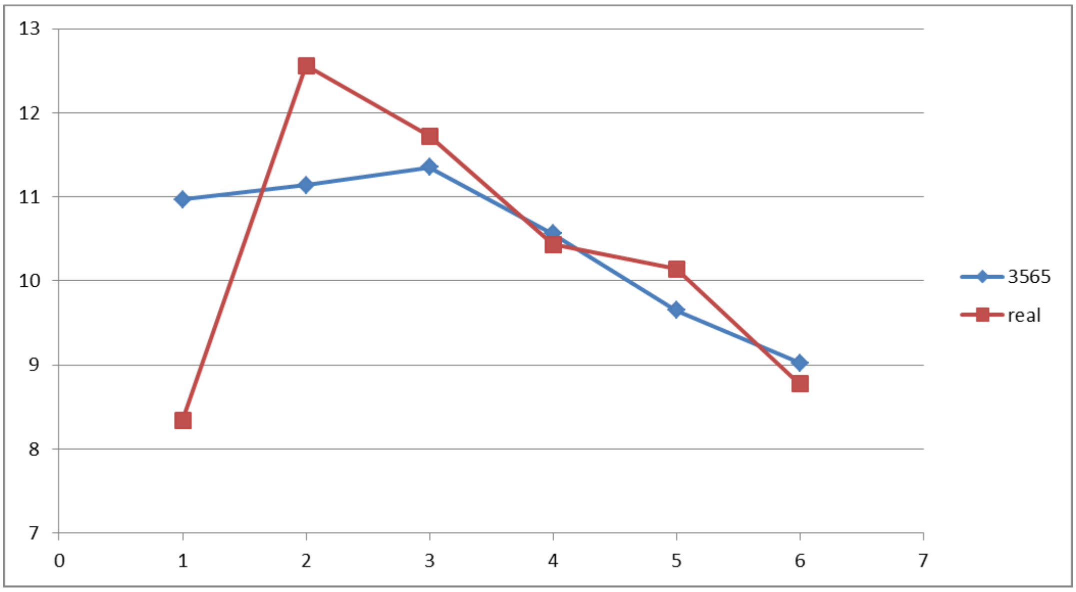

Figure 17. The real life data versus the simulation output for Brick 139

The graph in Figure 17 reveals that the trends of the results achieved in the simulation and most of the actual values are virtually the same. The starting market share is highly deviated due to used starting survey results which depict prescription habits (Rx Total and Rx Product) of customers.

Brick 627

The simulation has been run for the same parameter intervals with 20 repetitions of each run. This brick is particularly different because it has an odd detailing activity pattern the simulation has to execute, which varies highly from month to month and the distribution of calls to different customer segments also varies highly. The optimum parameters achieve for this brick is show in Table 19.

Table 19. Best Fit Parameters for Brick 627

Run number	4586
Alpha	1.2748
Memory A-C	0.17
Memory B-D-NoSeg	1.3536
Decay B-D-NoSeg	1.17
Decay A-C	0.88
[Final ms output]	0.98
Sum of errors squared	[5.23 5.26 5.19 4.92 4.82 5.27]

The results are similar to that of Brick 139. The memory effect for AC customers is 2.08 times higher and this is built into the simulation, but the decay factor for AC is also higher than B-D-NoSeg, pointing out to a similar finding that the doctors who do not prefer T-Pharma products have a longer carryover effect when called upon.

The sum of errors squared is significantly low; 1.27. If we calculated a similar percentage deviation for this brick, using the average market share of 5.05 and monthly error 0.46, we would arrive at an average of 9% deviation. Level of variation in the actual market share in this brick is not high, that is mainly why we can achieve a lower deviation in our predicted values compared to Brick 139. If we graph the predicted market share values from the simulation versus the real life values, we achieve the results as shown in Figure 18.

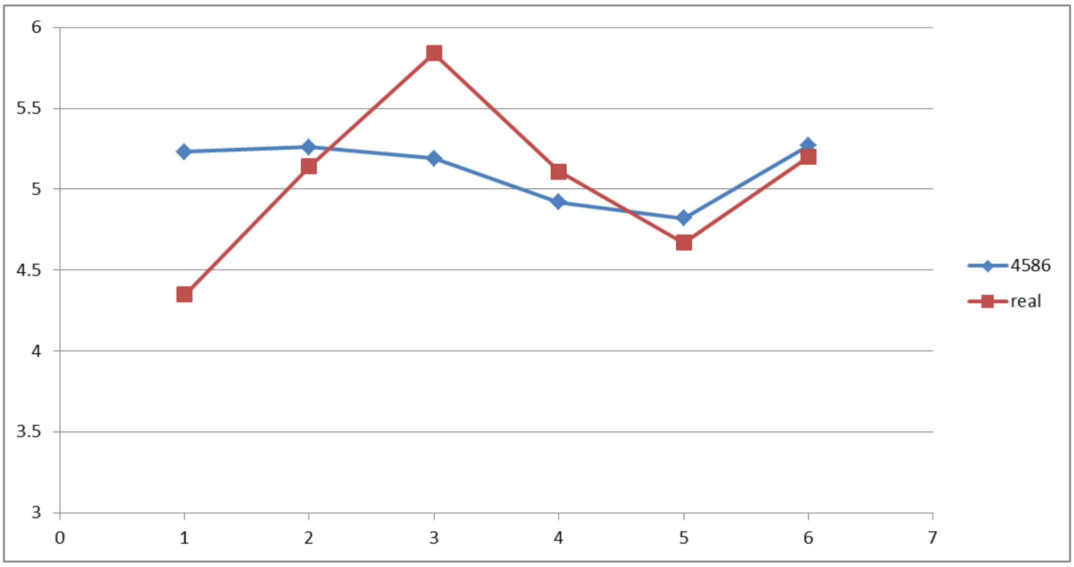

Figure 18. The real life data versus the simulation output for Brick 627

Again, the starting point is deviated due to inaccuracy of the survey data with the starting month's market share value, but the trend is virtually the same and the predicted values get closer to real life values, as months progress.

CHAPTER 5

CONCLUSION

This thesis focuses on the linear correlation between market share and detailing activities carried out by pharmaceutical companies and develops a simulation model that best predicts the real life data. The analyzed data is the actual activity and sales data for five products of a pharmaceutical company in Turkey. Three hypothesis have been tested against the data; in Hypothesis 1 the correlation and lag of detailing activities' effect on sales is determined, in Hypothesis 2 the effect of activities made to each customer segment (A, B, C, D, No-Segment) on sales is analyzed and in Hypothesis 3 the activities made to segments are grouped based on similar characteristics of customers and the relationship of the effect amongst them are determined. These findings are then used to develop a simulation model to find the optimum memory and carryover effects per different customer segments which best fit the actual values of market share.

Hypothesis 1 has revealed that a direct significant relationship exists for four out of five products. When the lag is analyzed, two of the products reveal a single month as the most explanatory independent variable while for the other two products, different months have similar significant correlation factors (i.e. $p = 1e\text{-}11$ and $p = 5e\text{-}11$). This is primarily due to the nature of the sales data available in Turkey and similar markets; showing sales figures from wholesalers to pharmacies at the month level while the lag may need to be measured in days and ideally should be measured from pharmacy to patient. Also, the correlations' R squared values are in a range that would be considered low; between 0.35 and 0.70. This is mainly because the detailing activities are not the only source of influence in changing the prescription preferences of physicians.

In Hypothesis 2, both R squared and t-statistics significantly decrease. Only Product-E has significant correlations for its visits made to four segments, while A and B had only two and C and D had none. The insignificant results may be caused by a faulty segmentation of the customers by the company or due to an inherent multicollinearity amongst the independent variables. Also, another important finding for the Product B and E is that the models reveal negative slopes for activities made to B, D or No-Segment customers. This is due to the *stealing effect* described in the thesis, where one visit made to these customer segments mean that a possible visit to the other A and C segments, where the impact is higher, is missed and thus the market share seems to fall when the sales representatives devote their time more on B, D and No-Segment customers than A and C customers.

Hypothesis 2 results lead to a test for multicollinearity between the independent variables and were found to be significantly collinear; VIF is found for all independent variables in the range of 6.33 - 10.89. This is due to total visits of each sales representative being similar every month and that visits to each segment are a portion within this overall sum. A good analogy is to think of visits to each segment as differently sized slices of the same pie. As each slice gets bigger, it must mean the others should get smaller. To overcome this, we have tried to apply partial regression coefficient finding methodology in which we would attain the net effect of each segment's visit, but that yielded statistically insignificant results with very low R squared values (in the range of 0.03 - 0.43). Lastly, we developed and tested Hypothesis 3 that coupled segments with similar characteristics to two independent variables; Visit-A-C and Visit-B-D-NoSegment. It was found significant (p = 1.4e-

14) and revealed the effect relationship between A-C and B-D-NoSegment customers.

The initial observation for the developed agent based model is that it is able to find optimum values within a range of 10-15% monthly deviation from the actual values for the selected bricks. The trend of the market shares in the actual data versus the best fit models match only when the fluctuations are little, within range of 20-25%, or when there is a single path trend, upwards or downwards. The randomized visits incorporated into the model effects the outcome of each month but the averaged market share outputs have a similar trend after 10 repetitions.

The simulation has predicted an average of 71% carryover effect for physicians in the selected bricks, whereas a similar study conducted in US had calculated an average of 74%, which had utilized a different model that employed neural networks and a nonlinear correlation (Yi et al., 2003). Another important difference is that Yi et al were able to map each sales figure with each physician and they did not have any lag in their correlation analysis. This thesis, although have used a linear method and a lagged sales data with no prescription data, it has attained similar results.

This thesis proposes a new methodology for the pharmaceutical industry for developing a solid statistical model, from which product and customer segmentation level correlation analysis may be attained and using these factors a simulation model may be build. This simulation model could be enhanced to include several bricks and several sales representatives for simulating effects of co-promotion and help in optimization of territory alignments. Also pharmaceutical companies may use this model to run different scenarios of visit planning, test their segmentation data quality and use it to forecast market share based on planned calls.

Brick 84

Table 20. Best Fit Parameters for Brick 84

Run number	1775
Alpha	17.0166
Memory A-C	0.06
Memory B-D-NoSeg	1.1248
Decay B-D-NoSeg	1.06
Decay A-C	0.98
[Final ms output]	0.98
Sum of errors squared	[10.49 9.78 9.66 9.5 8.99 9.49]

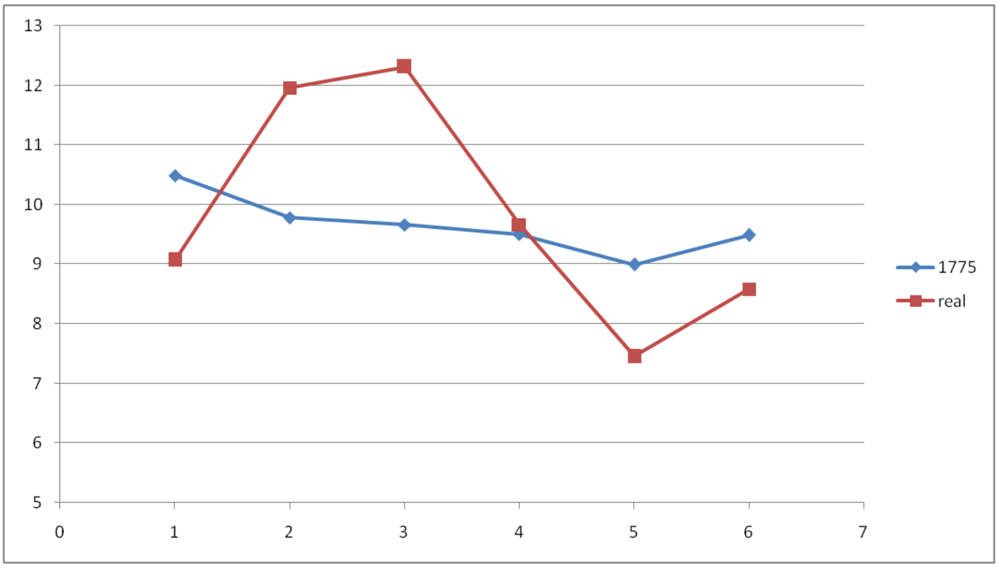

Figure 19. The real life data versus the simulation output for Brick 84

Brick 112

Table 21. Best Fit Parameters for Brick 112

Run number	6669
Alpha	10.5506
Memory A-C	0.26
Memory B-D-NoSeg	1.5408
Decay B-D-NoSeg	1.26
Decay A-C	0.94
[Final ms output]	0.7
Sum of errors squared	[13.7 12.97 12.64 13.26 12.55 11.62]

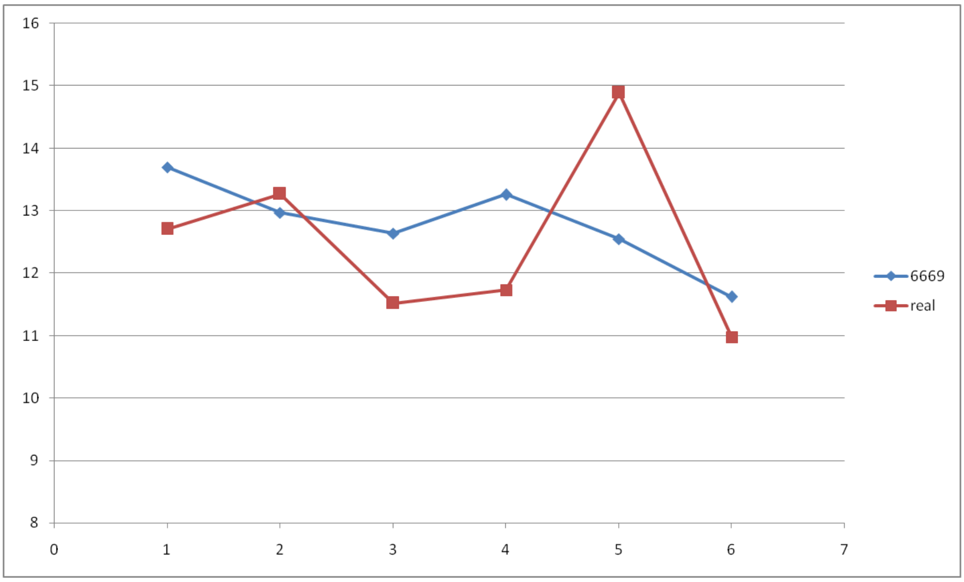

Figure 20. The real life data versus the simulation output for Brick 112

Brick 383

Table 22. Best Fit Parameters for Brick 383

Run number	3822
Alpha	9.5048
Memory A-C	0.14
Memory B-D-NoSeg	1.2912
Decay B-D-NoSeg	1.14
Decay A-C	0.96
[Final ms output]	0.98
Sum of errors squared	[9.88 10.1 10.15 9.97 9.85 9.59]

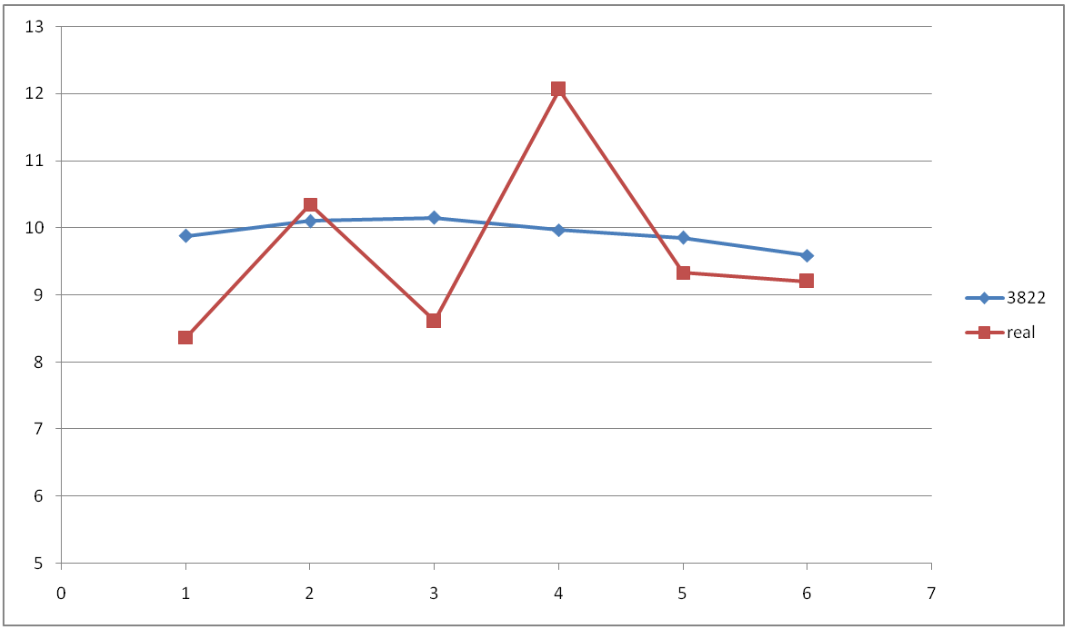

Figure 21. The real life data versus the simulation output for Brick 383

Brick 496

Table 23. Best Fit Parameters for Brick 496

Run number	12717
Alpha	396.353
Memory A-C	0.49
Memory B-D-NoSeg	2.0192
Decay B-D-NoSeg	1.49
Decay A-C	0.94
[Final ms output]	0.9
Sum of errors squared	[42.5 39.49 26.92 21.18 25.54 35.44]

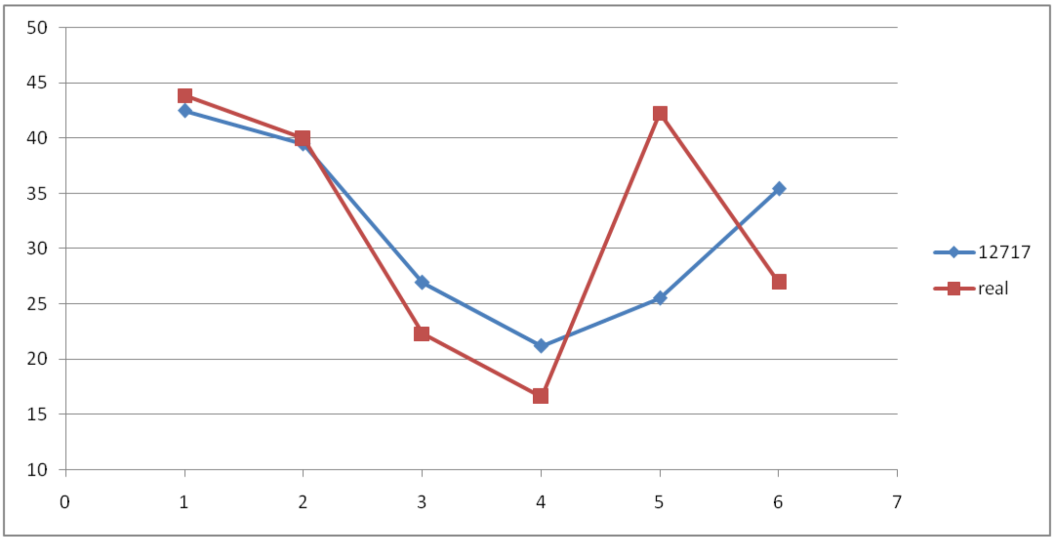

Figure 22. The real life data versus the simulation output for Brick 496

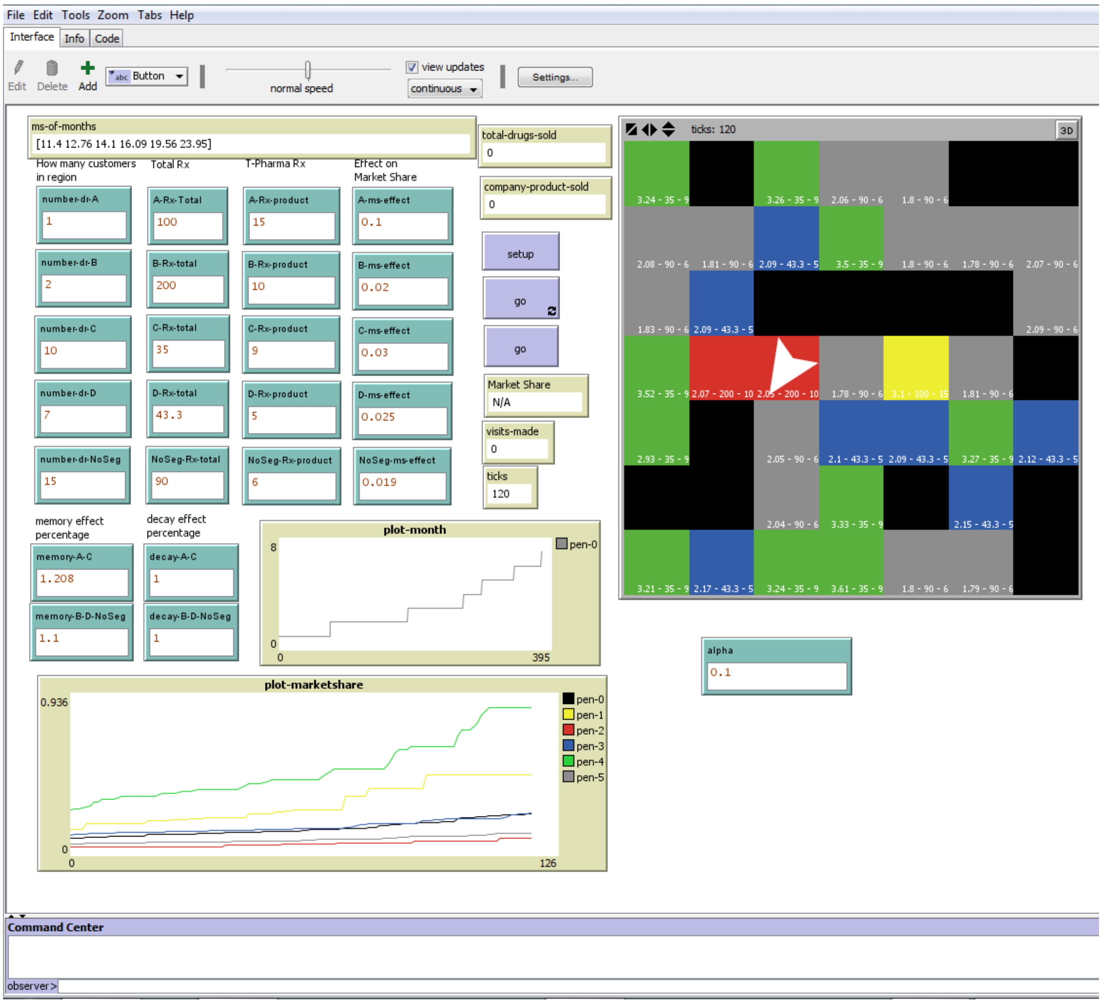

Figure 23. The user interface screenshot of the developed NetLogo simulation model

The script for the NetLogo simulation model developed

```
globals [
  doctors ;;the patch set which are doctors
  Total-visits-per-month


  visit-A
  visit-B
```

visit-C

visit-D

visit-NoSeg

visit-total

Target-A-freq

Target-B-freq

Target-C-freq

Target-D-freq

Target-NoSeg-freq

total-drugs-sold

company-product-sold

market-share

ms-of-months

has-started

rep

repetition

month ;; the current month, but it is -1 to make things easier for indexing

days-in-month ;;how many days each month has

```
  visits-made ;; to count how many visits have been done from the beginning of the
month
]

patches-own[
  segment ;;the segment of the doctor - A,B,C,D
  per-tick-Rx ;;how many prescriptions a doctor makes per tick
  per-tick-Rx-product ;;how many prescriptions for the company product the doctor
makes per tick

  total-products
  visit-at-freq ;;the frequency at which the doctor should be visited depending on
monthly available calls
  visit-cnt ;;how many times have this particular doctor been visited from the
begining of time
  last-visited  ;;the tick when this dr was last visited
  memory ;; the memory multiplier which decreases over time when the doctor is not
visited
  min-Rx-Product ;;the minimum number of company products a doctor would
prescribe (it is percentage ratio)
  memory-effect ;; the multiplier of how much influence a single visit causes
  memory-decay ;; the multiplier for the decay of the memory

  visit-left ;; this is the variable to see how many visits left for that customer segment
  ]
```

```
to set-global-variables
  set month 0
  set days-in-month 20
  set visits-made 0

  set visit-A [1 4 4 2 1 0]
  set visit-B [0 1 6 2 4 5]
  set visit-C [20 32 21 4 7 9]
  set visit-D [36 44 28 11 20 13]
  set visit-NoSeg [18 32 21 8 15 13]
  set visit-total [75 113 80 27 47 40]

  set Total-visits-per-month (item (month) visit-total)
  set Target-A-freq 3
  set Target-B-freq 3
  set Target-C-freq 3
  set Target-D-freq 3
  set Target-NoSeg-freq 3

  if has-started = 0 [set ms-of-months [0 0 0 0 0 0] set repetition 20 set rep 0] ;;this is
the variable to record all market shares of each month


  set total-drugs-sold 0
```

```
  set company-product-sold 0
end



to do-things-when-visited[a-patch]
  if segment = "A" [ set visit-A replace-item month visit-A ((item month visit-A) - 1)
]
  if segment = "B" [ set visit-B replace-item month visit-B ((item month visit-B) - 1) ]
  if segment = "C" [ set visit-C replace-item month visit-C ((item month visit-C) - 1) ]
  if segment = "D" [ set visit-D replace-item month visit-D ((item month visit-D) - 1)
]
  if segment = "NoSeg" [ set visit-NoSeg replace-item month visit-NoSeg ((item
month visit-NoSeg) - 1) ]

  ;;visit-at-freq
  ask a-patch [set memory (memory * (1 + ((memory-effect - 1) * ((ticks - last-
visited) / visit-at-freq)))) set last-visited ticks]
end

to go
  set has-started 1

  ;;output-print ms-of-months
  if month > 5 [reset-things-for-new-round set month 0 set repetition (repetition - 1)
set rep (rep + 1)]
```

```
  if repetition = 0 [
    set ms-of-months replace-item 0 ms-of-months precision (item 0 ms-of-months /
rep) 2
    set ms-of-months replace-item 1 ms-of-months precision (item 1 ms-of-months /
rep) 2
    set ms-of-months replace-item 2 ms-of-months precision (item 2 ms-of-months /
rep) 2
    set ms-of-months replace-item 3 ms-of-months precision (item 3 ms-of-months /
rep) 2
    set ms-of-months replace-item 4 ms-of-months precision (item 4 ms-of-months /
rep) 2
    set ms-of-months replace-item 5 ms-of-months precision (item 5 ms-of-months /
rep) 2
    stop
    ]

  set-doctors-visit-left
  rep-visit-doctor

  set visits-made visits-made + 1
  pass-day

  ask doctors [set plabel (word (precision memory 2) (word " - " (precision per-tick-
Rx 2)) (word " - " (precision per-tick-Rx-product 2)))]
```

```
  update-month
  update-month-plot
end

to pass-day
  if (visits-made >= ((item month visit-total) / days-in-month) * ((remainder (ticks)
days-in-month) + 1) )
  [
   ;;ask doctors [set per-tick-Rx-product per-tick-Rx-product * memory]
   ask doctors [prescribe-drugs]
   ;;ask doctors with [last-visited < ticks] [set memory memory * memory-decay]
   ask doctors with [last-visited < ticks] [set memory (memory * (1 - ((1 - memory-
decay) * (remainder (((ticks - last-visited) / visit-at-freq) - 0.0001) 1))))
     ;;output-print (word "decay factor: " (remainder (((ticks - last-visited) / visit-at-
freq) - 0.0001) 1) (word "memory carpan" ((1 - ((1 - memory-decay) * (remainder
(((ticks - last-visited) / visit-at-freq) - 0.0001) 1))))))
     ]

   update-marketshare-plot
   tick
   ]  ;;first update how many they will prescribe, then doctors prescribe then tick
end

to update-marketshare-plot
  set market-share company-product-sold / total-drugs-sold
```

```
  set-current-plot "plot-marketshare"

  set-current-plot-pen "pen-0"
  plot (market-share)

  set-current-plot-pen "pen-1"
  plot (sum ([per-tick-Rx-product * memory] of doctors with [Segment = "A"]) / sum
([per-tick-Rx] of doctors with [Segment = "A"]) )

  set-current-plot-pen "pen-2"
  plot (sum ([per-tick-Rx-product * memory] of doctors with [Segment = "B"]) / sum
([per-tick-Rx] of doctors with [Segment = "B"]) )

  set-current-plot-pen "pen-4"
  plot (sum ([per-tick-Rx-product * memory] of doctors with [Segment = "C"]) / sum
([per-tick-Rx] of doctors with [Segment = "C"]) )

  set-current-plot-pen "pen-3"
  plot (sum ([per-tick-Rx-product * memory] of doctors with [Segment = "D"]) / sum
([per-tick-Rx] of doctors with [Segment = "D"]) )
  set-current-plot-pen "pen-5"
  plot (sum ([per-tick-Rx-product * memory] of doctors with [Segment = "NoSeg"]) /
sum ([per-tick-Rx] of doctors with [Segment = "NoSeg"]) )
end
to prescribe-drugs
```

```
  set total-drugs-sold total-drugs-sold + per-tick-Rx
  set company-product-sold company-product-sold + (per-tick-Rx-product *
memory)

  ;;output-print (word "total-drugs-sold: " total-drugs-sold " company-product-sold:"
company-product-sold)
end

to update-month
  if ((item month visit-A) + (item month visit-B) + (item month visit-C) + (item
month visit-D) + (item month visit-NoSeg) = 0)
  [set ms-of-months replace-item month ms-of-months (item month ms-of-months +
precision (company-product-sold / total-drugs-sold * 100) 2)
   set total-drugs-sold 0
   set company-product-sold 0
   set month month + 1
   set visits-made 0
   if (month < 6)
     [
     set Total-visits-per-month (item (month) visit-total)
     ask patches with [segment = "A"] [set visit-at-freq floor Total-visits-per-month /
Target-A-freq]
     ask patches with [segment = "B"] [set visit-at-freq floor Total-visits-per-month /
Target-B-freq]
```

```
    ask patches with [segment = "C"] [set visit-at-freq floor Total-visits-per-month /
Target-C-freq]
    ask patches with [segment = "D"] [set visit-at-freq floor Total-visits-per-month /
Target-D-freq]
    ask patches with [segment = "NoSeg"] [set visit-at-freq floor Total-visits-per-
month / Target-NoSeg-freq]
    ]
  ]
end
to rep-visit-doctor
  ask turtle 0 [ move-to min-one-of (doctors with [visit-left > 0]) [visit-at-freq - (ticks
- last-visited)] ]
  ask turtle 0 [ do-things-when-visited patch-here]
end

to setup
  ca
  clear-output
  reset-ticks
  re-size-world ;; resize the world to show only entered number of doctors

  set memory-A-C 1 + (2.08 * alpha)
  set memory-B-D-NoSeg 1 + (1 * alpha)

  set-global-variables ;;set the visit pattern of the rep and other global variables
```

```
  create-doctors ;; create the patches representing each segment of doctors

  create-sales-rep ;, create a single turtle as a sales rep
end

to reset-things-for-new-round
  clear-patches
  clear-turtles
  clear-drawing
  clear-all-plots

  clear-output
  reset-ticks
  re-size-world ;; resize the world to show only entered number of doctors

  set memory-A-C 1 + (2.08 * alpha)
  set memory-B-D-NoSeg 1 + (1 * alpha)

  set-global-variables ;;set the visit pattern of the rep and other global variables
  create-doctors ;; create the patches representing each segment of doctors

  create-sales-rep ;, create a single turtle as a sales rep
end

to re-size-world
```

```
  let x ceiling sqrt (number-dr-A + number-dr-B + number-dr-C + number-dr-D +
number-dr-NoSeg) / 2
  resize-world  -1 * x x -1 * x x ;;only show optimum number of patches (doctors) in
the world
  set-patch-size 200 / x ;; normalize the size of the window depending on the number
of doctors
end

to create-doctors
  ask n-of number-dr-A patches with [pcolor = black] [ set pcolor yellow set segment
"A"]
  ask n-of number-dr-B patches with [pcolor = black]  [ set pcolor red set segment
"B" ]
  ask n-of number-dr-C patches with [pcolor = black]  [ set pcolor green set segment
"C" ]
  ask n-of number-dr-D patches with [pcolor = black]  [ set pcolor blue set segment
"D" ]
  ask n-of number-dr-NoSeg patches with [pcolor = black]  [ set pcolor grey set
segment "NoSeg" ]
  set doctors patch-set patches with [pcolor != black]

  set-doctors-visit-left ;; set remaining visits
  set-Rx-H-habits-of-doctors
  ask doctors [set memory 1 set last-visited -1 * visit-at-freq]
end
```

```
to set-doctors-visit-left
  ask doctors with [Segment = "A"] [set visit-left (item month visit-A)]
  ask doctors with [Segment = "B"] [set visit-left (item month visit-B)]
  ask doctors with [Segment = "C"] [set visit-left (item month visit-C)]
  ask doctors with [Segment = "D"] [set visit-left (item month visit-D)]
  ask doctors with [Segment = "NoSeg"] [set visit-left (item month visit-NoSeg)]
end

to create-sales-rep
  crt 1 [set color white]
end

to update-month-plot
  set-current-plot "plot-month"
  set-current-plot-pen "pen-0"
  plot (month + 1)
end

to set-Rx-H-habits-of-doctors
  ask patches with [segment = "A"] [set visit-at-freq floor Total-visits-per-month /
Target-A-freq set memory-effect memory-A-C set memory-decay decay-A-C set
per-tick-Rx A-Rx-total set per-tick-Rx-product A-Rx-product]
```

ask patches with [segment = "B"] [set visit-at-freq floor Total-visits-per-month / Target-B-freq set memory-effect memory-B-D-NoSeg set memory-decay decay-B-D-NoSeg set per-tick-Rx B-Rx-total set per-tick-Rx-product B-Rx-product]

ask patches with [segment = "C"] [set visit-at-freq floor Total-visits-per-month / Target-C-freq set memory-effect memory-A-C set memory-decay decay-A-C set per-tick-Rx C-Rx-total set per-tick-Rx-product C-Rx-product]

ask patches with [segment = "D"] [set visit-at-freq floor Total-visits-per-month / Target-D-freq set memory-effect memory-B-D-NoSeg set memory-decay decay-B-D-NoSeg set per-tick-Rx D-Rx-total set per-tick-Rx-product D-Rx-product]

ask patches with [segment = "NoSeg"] [set visit-at-freq floor Total-visits-per-month / Target-NoSeg-freq set memory-effect memory-B-D-NoSeg set memory-decay decay-B-D-NoSeg set per-tick-Rx NoSeg-Rx-total set per-tick-Rx-product NoSeg-Rx-product]

end

REFERENCES

Agnetis, A., Messina, E., & Pranzo, M. (2010). Call planning in European pharmaceutical sales force management. *IMA Journal of Management Mathematics*, *21*(3), 267–280. doi:10.1093/imaman/dpp019

Al-Hamdi, a. Y., Hassali, M. a. a., & Ibrahim, M. I. M. (2012). Impact of pharmaceutical promotion on healthcare professional's practices and behaviour: Views from general practitioners, medicine dispensers and medical representatives in Yemen. *Journal of Medical Marketing: Device, Diagnostic and Pharmaceutical Marketing*, *12*(4), 240–246. doi:10.1177/1745790412455942

Baier, M., Carballo, J., & Chang, A. (2012). Sales-force performance analytics and optimization. *IBM Journal of …*, *56*(6), 1–10. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6353942

Bates, a, Bailey, E., & Rajyaguru, I. (2002). Why pharmaceutical marketers must measure return on investment to ensure profitable eDetailing campaigns. *Journal of Medical Marketing*, *2*(4), 287–292. doi:10.1057/palgrave.jmm.5040087

Blythe, J. (2008). *Essentials of Marketing* (4th editio., p. 32). Pearson Education.

Central, P. (1998). COSTA : Contribution Optimizing Sales Territory Alignment : Abstract.

Drexl, A.; Haase, K. (1999). Fast approximation methods for sales force deployment. *Management Science1*, (45), 1307 – 1323.

Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology*, *58*(2), 107–12. doi:10.1016/j.jclinepi.2004.09.002

Gönül, F., Carter, F., Petrova, E., & Srinivasan, K. (2001). Promotion of prescription drugs and its impact on physicians' choice behavior. *Journal of Marketing*, *65*(July), 79–90. Retrieved from http://journals.ama.org/doi/abs/10.1509/jmkg.65.3.79.18329

Ihaka, R and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics, 5,* 299–314.

Kotler, P. (2000). *Marketing management: Millenium edition*. Englewood Cliffs: NJ: Prentice-Hall.

Lerer, L. (2002). Pharmaceutical marketing segmentation in the age of the Internet. *Journal of Medical Marketing*, *2*(2), 159–166. doi:10.1057/palgrave.jmm.5040069

Lodish, L. (1975). Sales Territory Alignment to Maximize Profit. *Journal of Marketing Research (JMR)*, (February), 30–37.

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162. doi:10.1057/jos.2010.3

Manchanda, P., Wittink, D. R., Ching, A., Cleanthous, P., Ding, M., Dong, X. J., … Xie, Y. (2005). Understanding Firm, Physician and Consumer Choice Behavior in the Pharmaceutical Industry. *Marketing Letters*, *16*(3-4), 293–308.

Puschmann, T., & Alt, R. (2001). Customer Relationship Management in the Pharmaceutical Industry, *00*(c), 1–9.

Taneja, G. (2008). Impact of Pharmaceutical Industry Promotion Mix on Doctor’s Prescribing Behaviour. *Asia-Pacific Journal of Management Research and Innovation*, *4*(4), 82–95. doi:10.1177/097324700800400408

Tokgöz, T. (2010). The Turkish pharmaceutical market. *Journal of Generic Medicines*, *7*(3), 270–274. doi:10.1057/jgm.2010.19

Wilensky, U. (2013). NetLogo. Evanston, IL: Center for Connected Learning and Computer-Based Modeling.

Wood, G. W., Gumbhir, a. K., Anderson, J. L., & Anderson, S. C. (1992). Strategic Planning in the US Pharmaceutical Industry: A National Study. *Therapeutic Innovation & Regulatory Science*, *26*(1), 63–76. doi:10.1177/009286159202600107

Wright, R. F., & Lundstrom, W. J. (2004). Physicians’ perceptions of pharmaceutical sales representatives: A model for analysing the customer relationship. *Journal of Medical Marketing*, *4*(1), 29–38. doi:10.1057/palgrave.jmm.5040140

Yi, J. C. (2008). An expert system to derive carryover effect for pharmaceutical sales detailing optimization. *Expert Systems with Applications*, *34*(3), 1742–1753. doi:10.1016/j.eswa.2007.01.037

Yi, J. C., “Anand” Anandalingam, G., & Sorrell, L. a. (2003). An expert system to physician-detailing planning. *Expert Systems with Applications*, *25*(4), 533–544. doi:10.1016/S0957-4174(03)00093-9

Zoltners, A. A. (2005). Territory Design : and of Modeling Implementation. *Marketing Science*, *24*(3), 313–331. doi:10.1287/mksc.l050.0133

Zoltners, A. A., & Sinha, P. (1983). Sales Territory Alignment: A Review and Model. *Management Science*, *29*(11), 1237–1256.