



# Cost-Effectiveness and Utility of Preemptive Pharmacogenomic Testing in Infants

## Citation

Khromykh, Alina. 2017. Cost-Effectiveness and Utility of Preemptive Pharmacogenomic Testing in Infants. Master's thesis, Harvard Extension School.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33825891>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Cost-Effectiveness and Utility of Preemptive Pharmacogenomic Testing in Infants

Alina Khromykh

A Thesis in the Field of Biotechnology Management  
for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2017



## Abstract

The goal of this work is to investigate whether preemptive pharmacogenetic testing offered in early childhood displays cost-efficiency and is an economically advantageous screening option to prevent adverse drug events. Pharmacogenomic testing has demonstrated clinical utility in many areas of healthcare. Yet due to the novelty of such tests, healthcare payers are reluctant to pay upfront for preemptive pharmacogenomic screening.

Cost efficiency of a clinical intervention has historically been measured by the incremental cost-effectiveness ratio expressed in quality adjusted life years or lived years – a comprise measure of the quality and quantity of life with and without the given intervention, defined by the difference in cost between two possible interventions, divided by the difference in their effect.

To date, only one study, which focused on adult patients and was conducted by O Alagoz et al, has evaluated the cost-effectiveness of one-time preemptive pharmacogenetic screening. Capitalizing on the data and principals described by O Alagoz et al, I developed a statistical predictive extrapolation model to assess cost-effectiveness for the early childhood group of patients. Additionally, other social factors such as household income and unemployment were evaluated for possible correlation. The outcomes of this case study demonstrated a significant cost-advantage of offering preemptive pharmacogenomic screening to the pediatric population. No conclusive

correlation between household income and unemployment rate with adverse drugs events was detected.

## Dedication

This effort is dedicated to Dr. Joe Vockley, who gave me courage to believe  
in my self-efficacy

## Acknowledgment

I'd like to thank the faculty and staff of the Harvard University Extension School for the exceptional opportunity to be their scholar. I wanted to recognize Maura McGlame and Steven Denkin for their compassionate dedication to my success.

I'd like to express my gratitude to the leadership (especially John Deeken) of Inova Health System's Inova Translational Medicine Institute. Without your continuous support, this thesis would not have been possible. I wanted to recognize my thesis director and mentor Dr. Ramaswamy Iyer, whose knowledge, guidance and encouragement were invaluable for this project. I wanted to thank my colleagues for their support and willingness to help, especially Julie Muskett, Allison Mitchel, Anatoly Ulyanov, Moin Ahmad, Natalie Hauser, Benjamin Solomon, Mehul Shah, and Thierry Vilboux. Additionally, I wanted to thank my dear friends and advisors Nathan Fosse, Oleksandr Koshchii, Alexander Kushnir, Madeline Holden and Liana Gorodetska for their help and support.

Special words of appreciation are extended to my family: Oleg and Yevgeniya Khromykh, Vladislava, Yevhen and Maria Koshchii as well as Laynel Feldshtein – their love and cheer kept me going.

Lastly, it is hard to express in words how grateful and blessed I am with the support of my husband, Gary Grinev – without his patience and love this work would not have been possible.

## Table of Contents

Dedication .....	iv
Acknowledgments.....	v
List of Tables .....	viii
List of Figures/Graphs.....	ix
I.      Introduction.....	1
Next-generation Sequencing.....	2
Personalized Medicine.....	6
Big Data Management.....	9
Pharmacogenomics.....	11
Inova Health System.....	13
II.     Materials and Methods.....	15
Incremented Cost-Effectiveness Ratio.....	15
Quality Adjusted Life Years.....	18
Simulation Model.....	20
Data Extrapolation.....	23
Least Squares Regression.....	23
Model Validation.....	24
Pharmacogenomic panel design and clinical pharmacogenetics guidelines.....	25

Unexpected Contributing Factors.....	27
III.    Results.....	29
Assessment of the given data.....	29
Data Extrapolation.....	30
Visualization of the data.....	31
Model Validation.....	32
Other Contributing Factors.....	34
Conclusion.....	37
IV.    Discussion.....	39
Bibliography.....	42

## List of Tables

Table 1. Incremental Cost Effectiveness Ratio for preemptive PGx in adults.....	2
Table 2. MediMap <sup>TM</sup> Pharmacogenomic panel at the Inova Health System.....	27
Table 3. ICER expressed in QALY and LY's.....	32

## List of Figures and Graphs

Figure 1. Sequencing Cost and Data Output Since 2000.....	7
Figure 2. Polymorphisms causing varying side effects .....	12
Figure 3. One Size Does Not Fit All: Pharmacogenomics in the realm of the Personalized Medicine.....	13
Figure 4. Incremental Cost Effectiveness ratio thresholds.....	18
Figure 5. Decision tree: represents the Markov model for genetic testing problem for a target age group. ....	21
Figure 6. Sample of the R code used to reorganize the dataset.....	29
Figure 7. Sample of the R code used to enhance data set .....	29
Figure 8. ICER data projections for adults.....	31
Figure 9: Aggregated outcomes of the ICER.....	33
Figure 10. Estimated annual Incidence of Adverse Drug.....	34
Figure 11. Unemployment rate vs percentage of discharged with ADEs.....	35
Figure 12. Household income vs. clinical encounters due to the ADE .....	36
Figure 13. Frequency and cost associated with occurrence of ADEs.....	37

## Chapter I

### Introduction

Advancement of the pharmaceutical industry in the past half-century has created many opportunities to treat, and in some cases cure, numerous medical conditions and diseases. Drugs are the prevailing commodities for the treatment of many sicknesses from the common cold to cancer. Despite the benefits of accurately prescribed medications and correct patient use, it has been noted that even properly administered drugs can have adverse effects. Adverse drug events (ADE) or reactions (ADR) are frequent complications of commonly prescribed drugs. They can include injuries resulting from the medical intervention related to the specific pharmaceutical agents and ADEs can range from mild symptoms such as a headache or local rash to life-threatening conditions such as anaphylaxis, syncope episodes, and even death (health.gov, 2016). In the inpatient care setting, ADEs accounted for an estimated one third of all hospital adverse events, affecting 2 million hospital stays each year and prolonging length of stay (LOS) at the in-patient wards by 1.7 to 4.6 days (health.gov, 11/17/2016).

Genetics is among the most novel factors that contribute to ADEs. Pharmacogenomics (PGx), an emerging field that studies how genes affect a person's response to drugs, enhances our understanding of the impact of genetic variation on drug

efficacy and toxicity. The goal of this work was to investigate if preemptive pharmacogenetic testing is cost-effective when offered earlier in life.

### Next-generation sequencing

By using genomic sequencing technology, PGx testing can detect variations in certain genes that code for proteins. These proteins, which drive drug responses, include a number of liver enzymes that convert medications from prodrugs into their active or inactive forms.

Technological advancement in the past decade has dramatically increased the accuracy of many genetic tests and decreased their costs due to a methodology called high-throughput massively parallel ‘next-generation’ genomic sequencing (Gallego et al., 2014; The European Bioinformatics Institute, 2016; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014).

These advancements are transforming the field of medicine, creating the opportunity to diagnose, predict, and prevent an ever-expanding number of medical conditions and diseases (Bodian et al., 2014). This has resulted in the development of many personalized health initiatives. For example, in his 2015 State of the Union Address, President Barack Obama announced the launching of his Precision Medicine Initiative. Thus, precision medicine has gained heightened recognition and publicity. (Collins & Varmus, 2015; [www.whitehouse.gov](http://www.whitehouse.gov), 2015).

Still, enormous headway in the fields of genetics and genomics must be gained before preemptive genomic sequencing tests will be widely offered and commonplace within the community medical setting. One area that demonstrates how genomic medicine is currently in active use (and holds even more potential) is PGx. Genomic

information from PGx testing can already be used to understand patients' responses to certain medications, enabling more tailored therapeutic choices in a variety of clinical contexts.

Awareness of PGx variants (genetic changes that can affect a gene's ability to regulate metabolism of the medications) that influence ADE risk and drug efficacy may inform prescribing decisions and reduce such events, as well as lower healthcare costs associated with suboptimal treatment outcomes.

However, detection, validation and justification of these variants are still a bottleneck (Zhao, Wang, Wang, Jia, & Zhao, 2013), as in many de-novo cases it is yet to be determined what is considered 'normal' and 'pathogenic' in genomic variation that range from single base changes to a large chromosomal-level alterations (Zhao et al., 2013).

One of the biggest obstacles in defining what is 'normal' and what is not is a lack of racial and ethnical diversity and the low statistical power of the genomic reference databases that are used as a comparing model for normal and abnormal aberrations in the genome (Bodian et al., 2015; Khromykh, A & Solomon BD, 2016). For example, if a genetic change is small or subtle, like a single nucleotide polymorphism (SNP) or a copy number variation of the segment of the DNA sequence that may be present in a certain proportion of the population, it is very hard to distinguish whether and to what extent such a change contributes to an observed phenotype. Detection and confirmation of *de novo* variants (new changes that have not been inherited from either parent, but are known from the literature and scientific databases to be proven disease-causing mutations) is relatively easy. Evaluation of novel variants (genetic changes that have

never been reported before, but seem to be pathogenic based on *in silico* models) often require additional work to achieve a satisfying explanation, including cell-based and animal models (such as zebra fish models, mice and others) (Bodian et al., 2015; Khromykh, A & Solomon BD, 2016).

One of the challenges to determining the nature of a genetic change is that laboratory-based endeavors are frequently time-consuming, limited in scope, and are not guaranteed to provide an unequivocal answer. Currently, publically available genomic reference databases, such as the Personal Genome Project (Price Ball, 2012) are lacking in broad racial, ethnic, and gender diversity, and computes genomic and clinical input primarily from middle-aged Caucasian men.

Another key issue involves controversies around privacy, ethics, and moral aspects of the research and clinical use of data from genomic sequences and related modalities. Current legislative protection of patient privacy raises great concerns for many stakeholders, as well as other related parties, stressing the need for a discussion on both personal and societal levels.

The psychosocial paradox of bias attitude towards the sharing of personal and health information, the lack of transparency at the State level, and the constant threat of an information breach all hold patients back from voluntarily sharing their health data with researchers. Lack of easily comprehensible information about the intended use, storage and further destiny of shared personal, clinical and genomic information may create a stressful environment for patients. This could be a major obstacle for the implementation of the PGx and other gene sequencing-based tests.

This case study will explore obstacles and opportunities of implementing

preemptive PGx testing in a large community-based health system, focusing on the cost-effectiveness of the test.

Currently in the United States, several healthcare systems have launched pilot programs implementing preemptive PGx testing that targets different patient age-groups (pediatric, adult, geriatric). Reports in the literature provide only limited information about implementation burden and contributing decision-making factors in the adoption of preemptive PGx testing at these institutions. (Dunnenberger et al., 2015) (Brixner et al., 2016; Johnson et al., 2013) (Owens, 1998).

This case study is designed to evaluate the contributing factors in aggregate to determine the feasibility of adopting preemptive PGx testing in the neonatal population specifically to study if there is a socio-economic advantage of offering preemptive PGx screening earlier in life. Cost-benefit will be evaluated using Incremental Cost-Effectiveness Ratio (ICER) (Azimi & Welch, 1998) (Owens, 1998) (Generics and Biosimilars Initiative (GaBI), 03/12/2010) expressed in the dollar value of the life-year gained with appropriate adjustments for quality of life, usually expressed as the quality-adjusted life-year ( QALYs) (Owens, 1998). In addition, this project will determine the baseline cost burden of ADE in pediatric and adult populations by evaluating the ‘Hospital Charge Data’ – a dataset of the 100 most common inpatient and 30 most common outpatient services. (data.gov, 2016). Lastly, this project will review current ethical, legislative and data-security concerns in genetic/genomic testing. Furthermore, it is my hypothesis that preemptive PGx testing is economically-justifiable and ethically appropriate in the neonatal population. In addition to showing the economic advantage of preemptive PGx testing, this case study will also discuss the hybrid value (including cost

advantage, improved quality of life, and decreased healthcare cost) of this type of testing, as health care stakeholders have myriad, often conflicting goals, including access to services, profitability, high quality, cost containment, safety, convenience and patient-centeredness (Porter, 2010).

### Personalized Medicine

Personalization of healthcare based on an individual's risk for disease and likelihood of therapeutic response bears great promise in improving patient care and lowering medical costs. Specifically, genomic sequencing and related modalities are transforming the field of medicine, creating the opportunity to diagnose, predict and prevent an ever-expanding number of medical conditions and diseases. A "precise approach" in relation to patient care refers to the tailoring of medical treatments to the individual characteristics of each patient. Personalized medicine is a highly anticipated consequence of the recent technological and scientific advances in the field of genetics and genomics, which builds upon the initial completion of the sequencing of the human genome (Lander et al., 2001).

The journey of this scientific breakthrough started over two decades ago. Led by the United States, the Human Genome Project was an international effort to sequence the entire human genome. This project, which took countless hours of human labor and cost over three billion dollars (Figure 1), utilized first generation sequencing techniques (i.e. Sanger's method) (Pavlopoulos et al., 2013). The impact of this event on the fields of biology, medicine, physiology and others is hard to overestimate.

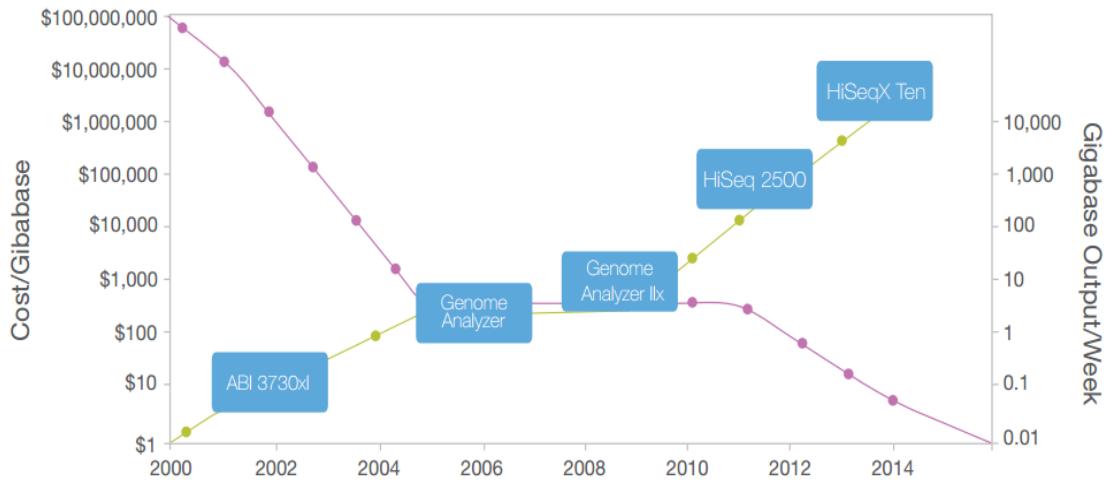


Figure 1. Sequencing Cost and Data Output Since 2000

Adopted from An Introduction to Next Generation Sequencing Technology, Illumina (An introduction to next-generation sequencing technology .2014; Illumina.2014; Raczy et al., 2013)

Many view Next-generation sequencing (NGS) as the most powerful diagnostic tool since the roentgenogram (Hennekam & Biesecker, 2012). However, it would be premature to call this technology a ‘crystal ball’ of medicine. Enormous work has to be accomplished before genomic sequencing will be more commonly used outside of the medical genetics setting such as in hospitals and in the offices of internists and pediatricians. Although the ability to generate tremendous amounts of genomic data is hard to resist, we must acknowledge the challenges, limitations, and ‘growing pains’ of this new diagnostic approach (Solomon, 2014).

One area where genomic medicine is already in active use and where there is considerable potential is pharmacogenetics. Information from PGx testing can be used to

understand patients' responses to certain medications, enabling more tailored therapeutic choices in a variety of clinical contexts (VanderVaart et al., 2011).

There is no doubt that the advancement of genomic technology, especially in the last five years has been astounding (Solomon, 2014). Widely applied, a second or a 'next generation' sequencing technology includes a number of different platforms (i.e. Illumina, Roche, etc.) that are able to generate 'raw' genomic data in a quick and relatively inexpensive manner (Amarasinghe, Li, & Halgamuge, 2013). However, analysis and interpretation of the vast amount of the data generated is not an easy or trivial task, and still remains an 'Achilles heel' of genomic and bioinformatics sciences (Pavlopoulos et al., 2013). Critics argue that although we are able to generate large amounts of data quickly, the ability to understand and handle this data is lacking (Solomon, 2014). For example, researchers using WGS to evaluate variants may be faced with 3-5 million candidate variants, a typical amount found in a single genome. Ideally, the number of candidate variants would be in the single digits where experts could more easily identify and validate true pathogenic mutations (Bodain et al., 2014)

It is yet to be determined what is considered 'normal' vs 'pathogenic' in genomic variation which can range from single base changes to large chromosomal-level alterations (Alkan, Coe, & Eichler, 2011; Zhao et al., 2013). To further clarify, a genomic variation can be a single nucleotide variation (SNV), or a small or large structural variation (SV) such as copy number variations that include insertions, deletions, duplications or translocation of genetic material. These changes do not necessarily have a negative impact on a person's health. Zhao et al, reports that approximately 12% of the human genome is subject to copy number variation (Zhao et al., 2013). However, the

presence of such changes could either be disease-causing, benign, or just unique to the particular individual, and creates challenges for accurate interpretation. Multiple methods including single-end mapping, pair-end approach, split-read mapping, utilization of a reference genome and others have been developed to address detection of structural variants. Countless software tools (i.e. BreakDancer, HyDra, CoNVE<sup>X</sup>, etc) have been implemented to assist scientists in their efforts to interpret structural variants, mostly focusing on the actual coding (exonic) regions of the genome that account for only 1% of the entire genome (Amarasinghe et al., 2013). Even though 85% of disease causing mutations are located in the coding regions, recent studies show that the non-coding portion of the genome contains equally important disease related information (Zhao et al., 2013). Special analytical bioinformatics software has been developed to detect these variants. To date, only a few such pipelines exist that can both cope with large amounts of data, efficiently analyze it, and detect genomic variation (Pavlopoulos et al., 2013).

Understanding of the ancestral and familial variants and their contributions to the personal genomic profile is essential for a well-rounded analysis. Such variants can be both informative and benign, but require additional information to be included in the reference genome, supporting the idea that not all reference genomes are created equally.

### Big data management

Another challenging area generated by genome sequencing is working with large amounts of data that is often measured in terabytes. The challenge is the ability to efficiently store and parse the data. Several file formats, such as FASTQ, SAM/BAM or VCF have been used to ease management of storing genomic data (Pavlopoulos et al.,

2013).

The quality and depth of sequencing the data plays an important role in the accuracy and reliability of the whole genome sequencing approach. Widely used, the Illumina HiSeq system can simultaneously generate several human genomes at ~30x coverage which delivers information in a cheaper and faster way (Pavlopoulos et al., 2013). However, some areas of the human genome require an additional in-depth investigation to determine further consequences of particular gene aberrations. For this purpose, researchers are utilizing methods of targeted resequencing (TR) for some portions of the genome (Zhao et al., 2013). The ability to magnify the depth of sequencing (six times or more) showed successful detection of some specific somatic mutations in the tissues of breast, ovarian, and prostate cancers (Zhao et al., 2013).

It is hard to question the ability of next generation sequencing to detect and discover many novel causes of human diseases, especially rare Mendelian disorders. However, molecular origins of complex and multifactorial conditions, such as obesity, diabetes, and pre-term birth are yet to be identified.

The detection and confirmation of ‘de-novo mutations’ (new disease-causing changes that have not been inherited from either parent) is a relatively easy task. However, identification of novel variants (those gene aberrations that have never been reported before but are believed to be disease-causing) may require additional work, including validation by in-silico and animal models (such as zebra fish models, mice and others).

Although the discussion on advantages and drawbacks of next generation sequencing is on-going, it is clear that further technological and analytical improvements

are essential to make genomic sequencing a cost and time efficient diagnostic tool. The real potential of data generated by next generation sequencing is yet to be fully comprehended. However, it is likely that in the next several decades WGS will dramatically increase our understanding of both commonly occurring and rare conditions, and replace many of the molecular diagnostic and screening tools in use today.

### Pharmacogenomics

Pharmacogenomics is the study of how a patient's genes regulate metabolism of a specific drug (Figure 2). Most medications are designed and given to patients based on standard prescribing guidelines that may not take into consideration one's genetic make-up (National Cancer Institute, 2016) (Figure 3).

Patient genotypes are usually categorized into the following predicted phenotypes (Pirmohamed, 2001; Sheffield & Phillimore, 2009)

- Ultra-Rapid Metabolizer: Patients with substantially increased metabolic activity;
  - Extensive Metabolizer: Normal metabolic activity;
  - Intermediate Metabolizer: Patients with reduced metabolic activity; and
  - Poor Metabolizer: Patients with little to no functional metabolic activity
- (Pharmainfo.net, 2016).

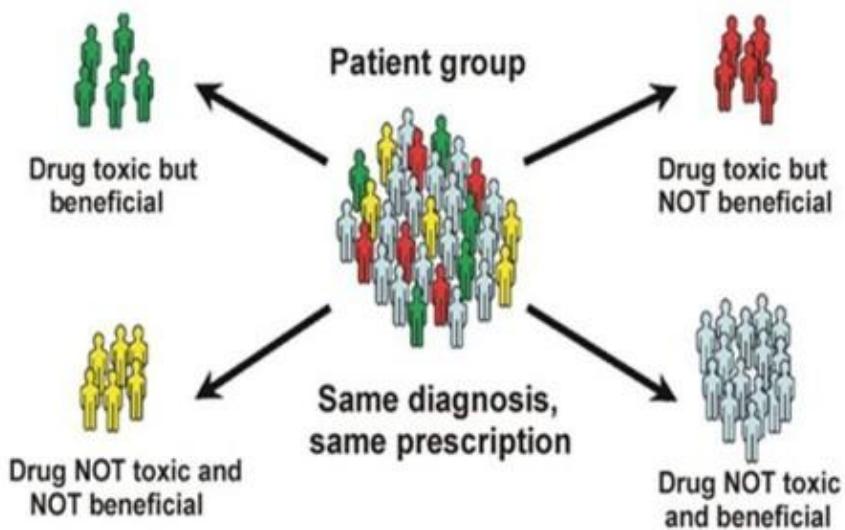


Figure 2. Polymorphisms causing varying side effects

Adopted from Pharmainfo.net: Polymorphisms causing varying side effects  
 (Pharmainfo.net, 2016)

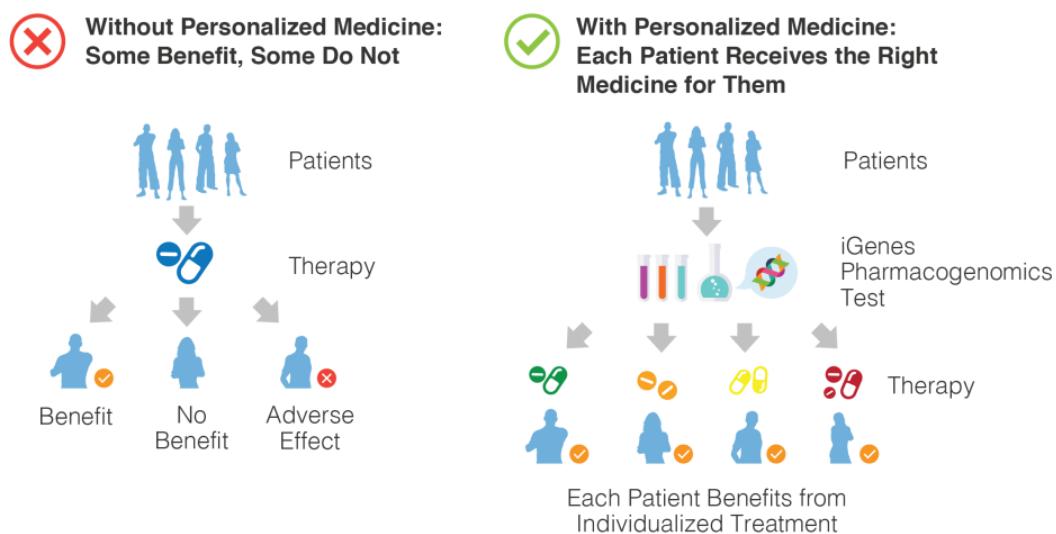


Figure 3. One Size Does Not Fit All: Pharmacogenomics in the realm of the Personalized Medicine

Adopted from Prenetics Medications: One Size Does Not Fit All (Prenetics, 2016)

## Inova Health System

The Inova Health System (IHS) and Inova Translational Medicine Institute (ITMI) provide one example of successful integration of genomics into clinical care. In one of its research studies, ITMI incorporates trio-based (Reed, 2014) whole genome sequencing data, input from electronic medical records, and additional information on stress, nutrition and environmental exposures obtained via study-specific surveys. The institute endeavors to catalog and correlate phenotype to genotype input from their participants who represent over 100 countries of origins (Hazrati, 2015). Another goal of this study is to enhance the understanding of less studied areas of genomic science, such as accurate minor allele frequencies, epigenetic interactions, and non-coding contributions to disease.

IHS is a non-for-profit community based health system with five hospitals and wide network of outpatient facilities. Inova services over 2 million patients a year. Specifically The Inova Fairfax Medical Campus (IFMC) is the system's largest medical center. It includes Inova Fairfax Hospital (IFH), an 876-bed tertiary care teaching hospital in Falls Church, Virginia, as well as Inova Women's Hospital (IWH) which delivers approximately 10,000 babies annually; the patient population is socioeconomically and racially and ethnically diverse.

Recently Inova Health System implemented an optional, preemptive PGx test, called MediMap<sup>TM</sup>, to all newborns delivered at IFH. (Inova.org, 2016; PR Newswire, 02/08/2016; Tracy Connell, 2016). MediMap is available on an opt-in basis to each and every newborn born at the IFH. Testing and analysis is performed at the Inova Genomics

Laboratory (IGL) and its contractual collaborator Translational Software. IGL is (CLIA)-certified clinical genomics laboratory.

## Chapter II

### Materials and Methods

Projections of the incremental cost-effectiveness ratio of the preemptive pharmacogenomic testing in early childhood expressed in quality-adjusted life year and lived years are outlined within this section. Extrapolative statistical modeling techniques were used for data projecting. All calculations were done utilizing statistical software (R-studio andXLSTAT ).

#### Incremented Cost-Effectiveness Ratio

Based on the dataset for the Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis In the United States (data.gov, 2016), hospitalizations due to ADE are included in the top 50 most common diagnoses .

Initial calculations reveal that the costs associated with the poisoning and toxic effects of drugs (with and without major complications) totals to ~\$15.3million for ~46 thousands of patient encounters (Attachment 1). Based on these statistics, we can calculate the average cost of a single patient encounter to be \$330.7 (\$15,272,688/46,177).

Current price for the PGx screening panels vary between \$300-\$2000 (NextGxDx). If we would offer preemptive PGx to the affected cohort (of ~46 thousands of patients) at the lowest price, it would be slightly cost efficient. However, this simple calculation doesn't factor in such important considerations that include quality-adjusted

life year and others. In contrast, the incremental cost-effectiveness ratio (ICER) is a common statistical analysis, used to summarize the cost-effectiveness of a health care intervention. This method is designed to calculate the difference in cost between two possible interventions, divided by the difference in their effect (Generics and Biosimilars Initiative (GaBI), 03/12/2010) (Azimi & Welch, 1998; Owens, 1998).

$$\frac{(\text{Cost with testing} - \text{Cost without testing})}{(\text{Health benefit with testing} - \text{Health Benefit without testing})}$$

(Owens, 1998)

The ICER compares two scenarios: presence of intervention (for example preemptive PGx test) with absence of it, to calculate cost-effectiveness as the difference in cost between two possible situations (with and without testing), divided by the difference in health benefit obtained (Owens, 1998) . In another words, this ratio represents as an extra payment per additional unit of effectiveness for more effective therapy.

ICER is a trusted approach to evaluate cost-effectiveness in healthcare economics (McCaffrey et al., 2015). Developed over 5 decades ago, the ICER has been utilized as a decision-making tool to justify fund allocation to pay for specific interventions. The ICER value has monetary expression and is evaluated against the set threshold.

Currently, literature reports a wide range of ICER thresholds across countries (M. Nanavaty, S. Kaura, M. Mwamburi, A. Gogate, J. Proach, A. Nyandege, Z.M. Khan, 2015) from \$13.000 (New Zealand) to \$104.000 (Canada) (M. Nanavaty, S. Kaura, M.

Mwamburi, A. Gogate, J. Proach, A. Nyandege, Z.M. Khan, 2015). Initially, in the United States the threshold per QALY was established at \$50.000. Some sources report that it has been raised to \$100.000 (Neumann, Cohen, & Weinstein, 2014)(Alagoz, Durham, & Kasirajan, 2016), however here we considered both threshold parameters.

ICER approach is not flawless, and certain controversies exist around it (William S. Weintraub, David J. Cohen, 2009). Some argue that basing decisions for or against health care intervention solely on the cost-effectiveness ratio may limit the number and type of interventions offered to patients (Oppong, Jowett, & Roberts, 2015). Therefore, other considerations should be included into computation when evaluating the adoption of a new intervention. However, it is deemed appropriate to utilize ICER for evaluation of the preemptive pharmacogenomic testing, as described in literature (Alagoz et al., 2016) .

O Algoz et al describes three primary approaches of economic evaluations in healthcare: cost-benefit, cost-utility and cost-effectiveness. In all three approaches, economists measure health outcomes. The difference between each, is the way that outcome is defined and valued.

In the cost-effectiveness analysis we focus on the one clinical effect of interest, such as number of adverse events prevented, number of life-years gained, and amount of funds saved. This analysis provides a measure of effectiveness based on the invested cost. Combining cost and utility, cost-utility analysis is a more feasible evaluation of laboratory tests. This type of analysis measures health outcomes in single and multiple effects in QALY from the diagnostic intervention.

The lower the QALY value is, the more economically efficient it is to offer a specific test or intervention (Figure 4).

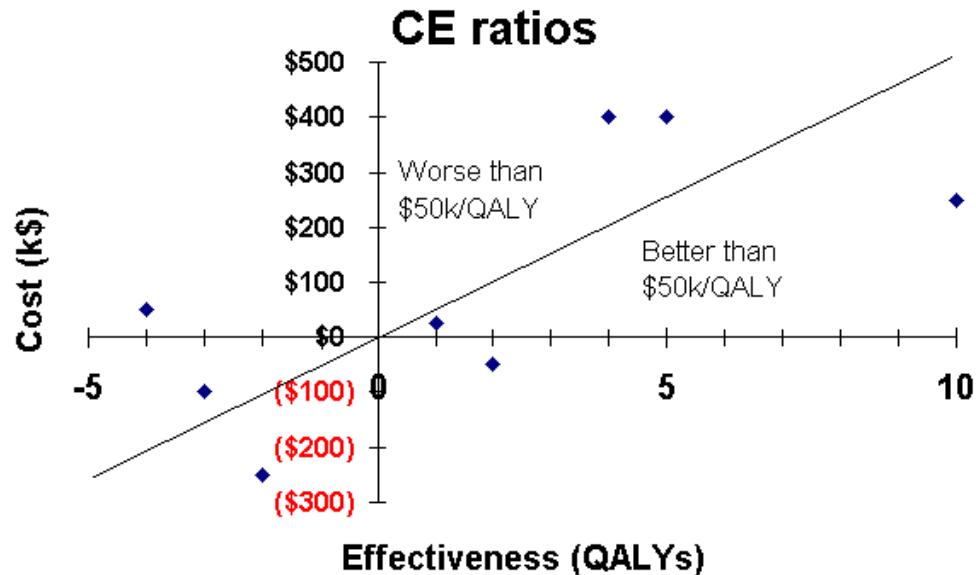


Figure 4. Incremental Cost Effectiveness ratio thresholds

Adopted from <http://araw.mede.uic.edu> (Alan Schwartz, 2001)

#### Quality Adjusted Life Years

Quality Adjusted Life Years (QALY) comprise measure of the quality and quantity of life with and without given intervention (National Institute for Health and Clinical Excellence, 2009). Specifically, O Alagoz et al describes the evaluation of the one-time genetic testing to minimize lifetime adverse drug reactions expressed in ICER. The measured health outcomes were expressed in life years (LYs) and quality-adjusted LYs (QALYs) for the cohort of adult patients (40-years old and up) (Alagoz et al., 2016). Here, QALY refers to a composite measure of the quantity and quality of life with and

without pharmacogenetic testing expressed in the dollar value. Inclusion of QALY into computation may provide a better appreciation of the overall health benefits, harms and costs of laboratory tests in the diagnostic decision making process and the induced health outcomes (Alagoz et al., 2016). Consideration of QALY measurement in computation will enhance the statistical predictive model. Specifically, the model includes projection of the potential benefit of preemptive testing not only from the absence of medical cost but also from prevented ADEs. Prevention of ADEs can ultimately improve longevity. Cost per year gained by the prevention of ADEs will be expressed via QALY in US dollars.

This case study aims to calculate ICER of preemptive PGx testing offered to the young pediatric population using a simulation statistical model to estimate the incremental cost-effectiveness of preemptive PGx testing compared with no genetic testing over the course of life (average Life expectancy 78.74 years) (Larry Copeland, 10/9/2014) (Alagoz et al., 2016). Similar to the study done by O Alagoz et al, only direct medical costs will be taken into consideration to avoid excessive confounding variables. Additionally, this analysis will consider the frequency of ADE in neonatal, pediatric and adult populations to further assess cost-effectiveness of the test if offered on the screening basis to the entire population

Extensive literature review didn't reveal that cost –effectiveness of preemptive multigene PGx screening was ever calculated for the neonatal population. This case study is designed to fill the gap in knowledge of cost-effectiveness and burden of ADE on health outcomes over the lifetime vs. cost of the preemptive PGx screening

with consideration of frequencies of ADEs.

### Simulation model

Predicting health outcomes from a hypothetical adverse drug event is not a trivial task. Advanced statistical data and capabilities are required to correctly establish base case scenario and all possible outcomes to build a valid model. To date, only one research study has achieved that. O Alagoz et al developed a Markov-based Monte Carlo model to estimate ICER of one time PGx testing versus no PGx testing for the 40 year old hypothetical patient (Alagoz et al., 2016). O Alagoz's model was developed based on three primary parameters:

- 1) ADR rate, composed of retrospectively collected stats of ADR-related encounters, which caused patient to seek medical attention (Emergency department, Outpatient care, Hospitalization).
- 2) Probable reduction in rate of occurrence of the ADR when genetic testing was used.
- 3) Estimated probability of death, caused by ADR.

The figure below (decision tree) schematically indicates the decision-making model (Figure 5).

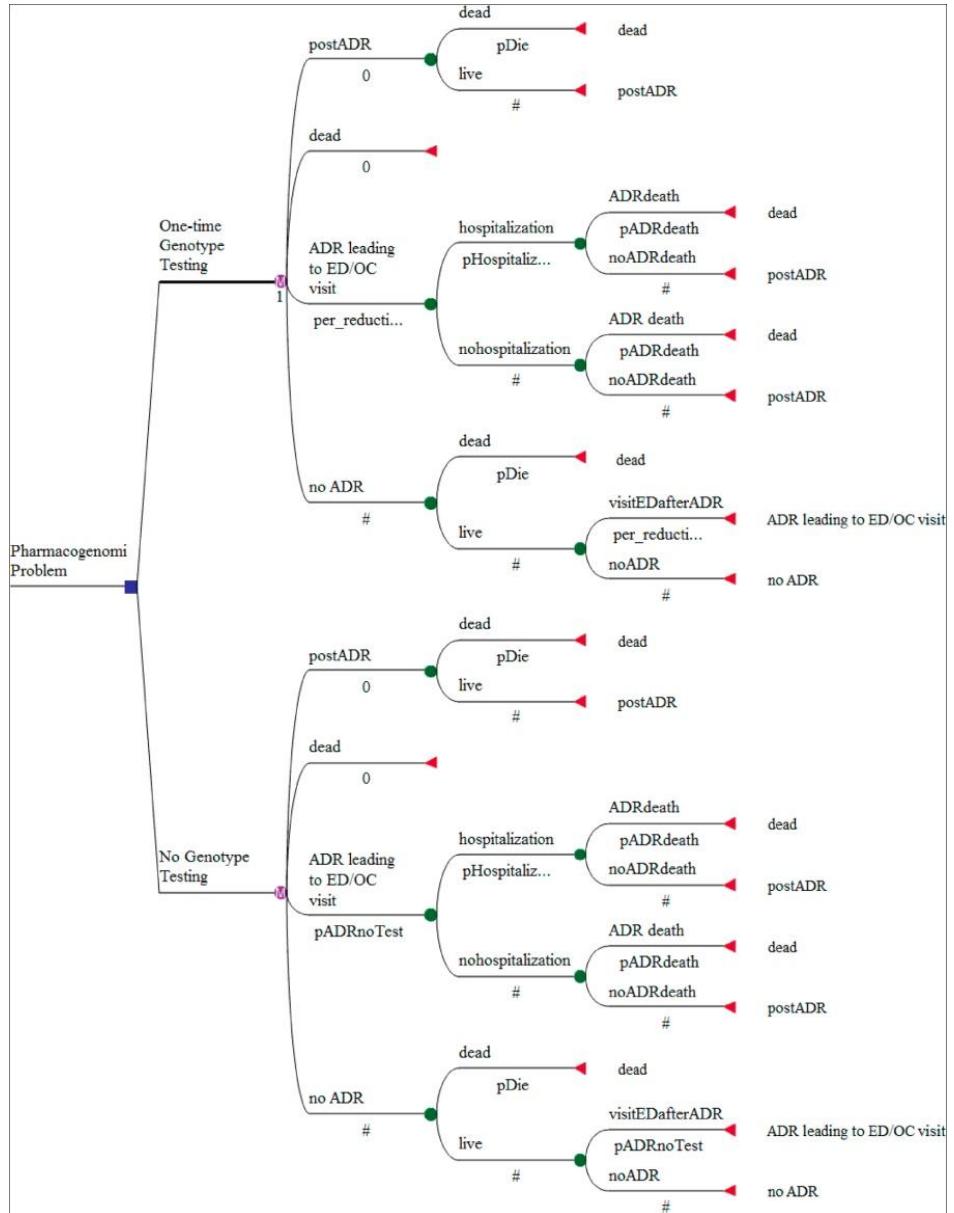


Figure 5. Decision tree: represents the Markov model for genetic testing problem for a target age group.

Note: This figure shows the conceptual model used in this study. In the figure, the square represents decision nodes, the circles represent chance nodes (random events), the reverse triangles represent the outcomes/end points and the node with the 'M' label shows the Markov nodes. A patient entering the simulation model is either offered a genetic testing or not. Then, at every age, the patient may experience an ADR that leads the patient to visit the emergency department (ED) or outpatient clinic (OC) with a probability estimated from the literature. Once the patient visits ED/OC due to an ADR, given certain probabilities the patient may be hospitalized or die.

Adopted form (Alagoz et al., 2016)

This model was used to measure health outcomes from the one-time genetic testing in life years (LY's), QALY's costs and calculated based on these values ICER, as cost per LY's and QALY gained. The calculations were developed for the following age groups, with outcomes respectively (Table 1):

Table 1: Incremental Cost Effectiveness Ratio for preemptive PGx in adults

Outcome table for the base-case results									
(a) Various clinical outcomes									
Target age group (i.e., if we screen this group)	Number of ED/OC visits per 1000 patients (no testing)	Number of ED/OC visits per 1000 patients (testing)	P-value	Number of hospitalizations per 1000 patients (no testing)	Number of hospitalizations per 1000 patients (testing)	P-value	Number of ADR deaths per 1000 patients (no testing)	Number of ADR deaths per 1000 patients (testing)	P-value
40	598	503	< 0.001	35	29	< 0.001	19	16	< 0.001
50	548	461	< 0.001	36	31	< 0.001	19	16	< 0.001
65	477	400	< 0.001	40	33	< 0.001	18	15	< 0.001
70	447	375	< 0.001	39	33	< 0.001	17	14	< 0.001
75	412	345	< 0.001	37	31	< 0.001	17	14	< 0.001
(b) Cost and life-year outcomes									
Target age group	Total expected discounted cost per person (no testing)	Total expected discounted cost per person (testing)	Total expected discounted life years per person (no testing)	Total expected discounted life years per person (testing)	Total expected discounted QALYs per person (no testing)	Total expected discounted QALYs per person (testing)	Cost-effectiveness ratio of genetic testing for this target group (per LYs)	Cost-effectiveness ratio of genetic testing for this target group (per QALYs)	
40	\$987	\$1930	23.1418	23.1636	19.4430	19.4605	\$43 165	\$53 680	
50	\$1068	\$1998	19.9978	20.0195	16.3839	16.4011	\$42 830	\$54 335	
65	\$1231	\$2134	14.3799	14.3992	11.3072	11.3219	\$46 892	\$61 465	
70	\$1247	\$2147	12.3281	12.3446	9.4719	9.4843	\$54 684	\$72 651	
75	\$1230	\$2132	10.2364	10.2511	7.7669	7.7778	\$61 440	\$82 632	

Abbreviations: ADR, adverse drug reaction; ED, emergency department; OC, outpatient clinic; QALY, quality-adjusted life year.

Note: Adopted from (Alagoz et al., 2016)

## Data Extrapolation

Extrapolation is a feasible statistical approach used to estimate values beyond the specific range of a given variable. Specifically, extrapolation allows estimation of the observation below the given values. This method assumes that the existing trend will continue and can be predicted mathematically. There are several methods of how to extrapolate data; the choice of method depends on the type of relationship contained in the data set. To choose between the linear, polynomial, conic and French curve – one must first hypothesize the nature of the trend (J. Scott Armstrong, Fred Collopy, 1993).

## Least Squares Regression

Scatterplot is a way of visualizing the data set in an attempt to assess a relationship between data points. If the plotting resembles a linear relationship pattern, this could be confirmed by drawing a line through the scatterplot (Yates). This line can be viewed as the “best” line for the purposes of describing and predicting a quantitative variable with an outcome. The line gives us an estimated mean of the outcome for each value of our predictor variable with only two parameters: the intercept and the slope. Simple linear regression models the expected mean of an outcome of one variable conditional on values of input variable. In simple linear regression, we predict scores of one variable from the scores of a second variable. Graphical visualization of the data set will help me to illustrate the presence or absence of correlation between given variables. The direction of a correlation is either positive or negative. In a negative correlation, the variables move in inverse, or opposite, directions. In other words, as one variable

increases, the other variable decreases.

The difference between the fitted values and actual values of the outcome is called the residuals (or error) term. The values of the intercept and slope from simple linear regression are obtained through the method of least squares (that is, by minimizing the SSE) By minimizing the SSE, the intercept, slope, the errors (or residuals), and the predicted values have several particular properties.

$$r = \frac{\sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

Where:  
 $\bar{x}$  is the sample mean of the predictor variable  
 $s_x$  is the sample standard deviation of the predictor variable  
 $\bar{y}$  is the sample mean of the response variable  
 $s_y$  is the sample standard deviation of the response variable  
 $n$  is the sample size

The Equation for the Least-Squares Regression

$$\hat{y} = b_1 x + b_0$$

where

$b_1 = r \cdot \frac{s_y}{s_x}$  is the slope of the least-squares regression line

and

$b_0 = \bar{y} - b_1 \bar{x}$  is the y-intercept of the least squares regression line. (Elgin Community College; Michael Sullivan, III, Joliet Junior College, 2012)

### Model validation

Data extrapolation model will be predicting ICER for pediatric and young-adult groups based on the output values of ICER generated by O Alagoz's et al. Type of data relationship (linear, non-linear, etc.) will dictate the appropriate calculation method. Preliminarily, the evidence suggests that the data set displays a positive linear relationship of data points (ICER per QALY and ICER per LY's). QALY and LY's are

aggregate functions of 3 average rates: ADR rate, composed of retrospectively collected stats of ADR-related encounters, which caused patient to seek medical attention; the probable reduction in rate of occurrence of the ADR when genetic testing was used and estimated probability of death, caused by ADR. Based on the design of these variables, we observe that the probable reduction in ADR due to the genetic testing and estimated probability of death based on ADR will remain -unchanged for any age group. Lastly, the ADR rate of clinical encounters due to unintended drug poisoning was comprised of the retrospectively collected data from medical centers.

To confirm the linearity of the relationship of the data points predicted for the pediatric and younger adults, we need to confirm that each component of these data point have a linear direction. That could be done by projecting actual percentages of reported cases of clinical encounters due to an ADR including emergency room visits, outpatient clinic encounters and hospitalization rates (Bates et al., 1997).

Pharmacogenomic panel design and clinical pharmacogenetics guidelines

Inova deferred to the PharmGKB and the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines to decide which gene-drug pairs and variants to include in our PGx panel. Priority was given to gene-drug pairs with clearly actionable evidence levels (versus prioritizing gene-drug pairs with lower evidence or clarity of use but which are more commonly prescribed to children). This selection process intentionally resulted in a small panel of 21 gene-drug pairs from among the variants available on the sequencing platform (Table 2). All gene-drug combinations

chosen have a “1A” level of evidence per PharmGKB (<https://www.pharmgkb.org/>) and implementation in the U.S. A level 1A PharmGKB distinction indicates a variant-drug combination that has been endorsed in the CPIC’s or other medical society’s guidelines or has been implemented at a Pharmacogenomics Research Network (PGRN) site or other major health system.

Table 2. MediMap™ Pharmacogenomic panel at the Inova Health System

Category	Therapeutic Class	Drug Name	Gene
AntiCancer Agents	Purine Antagonist	Mercaptopurine	TPMT
	Purine Analog	Thioguanine	TPMT
Cardiovascular	Anticoagulants	Warfarin	CYP2C9/VKORC1
	Antiplatelets	Clopidogrel	CYP2C19
	Statins	Simvastatin	SLCO1B1
Pain	Opioids	Codeine	CYP2D6
Psychotropic	Antidepressants	Amitriptyline	CYP2C19/CYP2D6
	Antidepressants	Citalopram	CYP2C19
	Antidepressants	Clomipramine	CYP2C19/CYP2D6
	Antidepressants	Desipramine	CYP2D6
	Antidepressants	Doxepin	CYP2C19/CYP2D6
	Antidepressants	Escitalopram	CYP2C19
	Antidepressants	Fluvoxamine	CYP2D6
	Antidepressants	Imipramine	CYP2C19/CYP2D6
	Antidepressants	Nortriptyline	CYP2D6
	Antidepressants	Paroxetine	CYP2D6
	Anticonvulsants	Phenytoin	CYP2C9
	Antidepressants	Sertraline	CYP2C19
	Antidepressants	Trimipramine	CYP2C19/CYP2D6
Immunologic	Immunosuppressants	Tacrolimus	CYP3A5
	Immunosuppressants	Azathioprine	TPMT

Note: Adopted from (Inova.org, 2016)

## Unexpected contributing factors

In addition to the primary contributing factors that I discussed earlier in this chapter, I hypothesized that there might be clinically unrelated and logically unexpected contributing factors which influence the rates of ADE. These factors might include, but are not limited to unemployment and household income.

To determine if there is a correlation between unemployment, household income and the frequency of ADRs, I reviewed the Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) –a dataset publically available via data.gov website (health.gov, 11/17/2016). The initial dataset contains n=163065 i=1 to n cases, where each case (i) represents a specific reason for seeking medical attention. Among these cases I will focus on the cases coded 917 and 918 that represent ‘Poisoning and toxic effect of drugs with and without major complications and comorbidities’, respectively.

The initial dataset includes 11 variables that are either categorical: Provider Id, Provider Name, Provider Street Address, Provider City, Provider State, Provider Zip Code, Hospital Referral Region Description or numerical: Total number of Discharges, Average Covered Charges, Average Total Payments, Average Medicare Payments.

The dataset has a granularity of cases based on the specific reason for seeking medical attention and I would like to review the instances of ADEs (cases coded 917/918) in each of the 50 states. In order to do that I have manipulated the data set to change the nature of the case rows. Utilizing R studio I have comprised cases by state and reorganized my data set to have n=50, where i=state (Figure 6).

```

40 #Creating 2 new tables called "v917" and "v918" by coping the corresponding set of data from our dataset called "mydata"
41 v917 <- mydata[mydata$DRG.Definition == '917',]
42 v918 <- mydata[mydata$DRG.Definition == '918',]
43
44 #To sum up numbers for each state, we first need to delete a dollar sign from our dataset
45 v917$Covered.Charges <- as.numeric(gsub("\\$", "", v917$Covered.Charges))
46 v917$Total.Payments <- as.numeric(gsub("\\$", "", v917$Total.Payments))
47 v917$Medicare.Payments <- as.numeric(gsub("\\$", "", v917$Medicare.Payments))
48
49 #now we can sum up the corresponding numbers
50 v917 <- ddply(v917, "State", transform, Discharges=sum(Discharges), Covered.charges=sum(Covered.charges),
51 Total.Payments=sum(Total.Payments), Medicare.Payments=sum(Medicare.Payments))
52
53 #here we delete rows that are repeating
54 v917 <- subset(v917, !duplicated(state))
55
56 #Fill in the column "Covered.Charges.Per.Person" with the numbers that we estimate by formula
57 #'Covered.Charges.Per.Person' = "Covered.Charges" / "Discharges"
58 v917$Covered.Charges.Per.Person <- v917$Covered.Charges / v917$Discharges
59 v917$Total.Payments.Per.Person <- v917$Total.Payments / v917$Discharges
60 v917$Medicare.Payments.Per.Person <- v917$Medicare.Payments / v917$Discharges
61

```

Figure 6. Sample of the R code used to reorganize the dataset

Next, I have supplemented my data set with additional numerical variables: average income per state and average unemployment rate per state. The enhancement of the data set with these additional variables allowed me to develop simple linear regression models and investigate the potential correlation between the instances of ADEs and unemployment rates, as well as ADEs' cost and the average income for the given state (Figure 7).

```

61
62 population = read.csv("population.csv")
63 colnames(population)[2] <- "Population"
64 v917$Population <- population$Population
65
66 income = read.csv("income.csv")
67 colnames(income)[3] <- "Income"
68 v917$Income <- income$Income
69
70 unemployment = read.csv("unemployment.csv")
71 colnames(unemployment)[2] <- "Unemployment"
72 v917$`Unemployment` <- unemployment$Unemployment

```

Figure 7. Sample of the R code used to enhance data set

## Chapter III

### Results

Scatter-plot visualizations based on the data extrapolation approach were successfully created. The projected incremental cost-effectiveness ratio of the preemptive pharmacogenomic testing in early childhood was calculated and expressed in quality-adjusted life years and lived years. Data collected from these calculations are thoroughly described in this section.

#### Assessment of the given data

To determine the relationship in the data-outcomes developed by O Alagoz et al, I have developed a plot with data points given in the article (Figure 8). To assess the linearity of the relationship, I have calculated the  $R^2$  for each series of data points reflected on the plot.  $R^2$ , which ranges between 0 and 1 (1= 100%), is a recognized statistical measure of the proximity of a given data unit to the fitted line on the plot. The closer the value of  $R^2$  to 1 the better proposed model represents our data set.

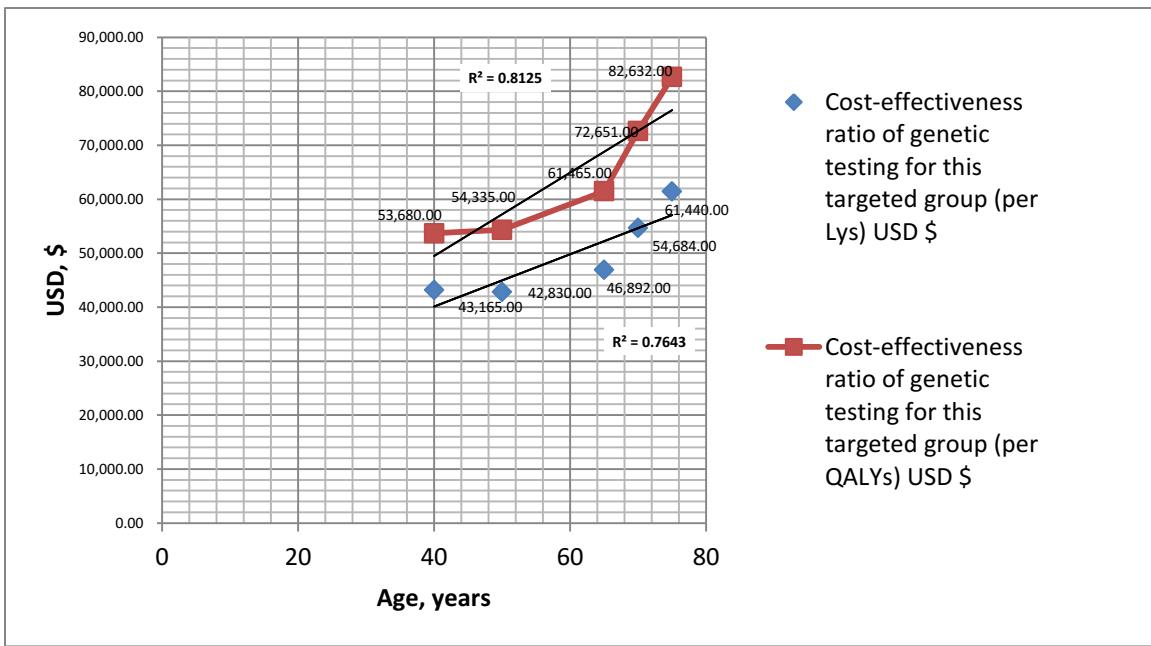


Figure 8. ICER data projections for adults  
Developed based on the output data from (Alagoz et al., 2016)

#### Data extrapolation

Based on the detected linear relationship of the cost-effectiveness ratio to patient's age, I have calculated the projected values of the ICER expressed in QALY and LY utilizing the least squares linear regression equation (Table 3).

Table 3: ICER expressed in QALY and LY's

Age	Cost-effectiveness ratio of genetic testing for this targeted group (per LY's) USD \$	Cost-effectiveness ratio of genetic testing for this targeted group (per QALYs) USD \$
Predicted value	5 23188.4	22481.95
	10 25607.8	26342.9
	20 30446.6	34064.8
	30 35285.4	41786.7
Existing value	40 43,165.00	53,680.00
	50 42,830.00	54,335.00
	65 46,892.00	61,465.00
	70 54,684.00	72,651.00
75	61,440.00	82,632.00

Note: Values for the age groups of 40, 50, 65, 70 and 75 years were adopted from O Alagoz et al (Alagoz et al., 2016).

Values for the age groups of 5, 10, 20 and 30 years were calculated by least squares linear regression equation.

#### Visualization of the data

The predicted and adopted values were projected on the plot to visualize the linearity of the relationship between the data points (Figure 9).  $R^2$  was calculated for each series of data points to evaluate how close the data points were to the fitted regression line. The  $R^2$  value of ICER expressed in LY's was  $R^2=0.9543$ , the  $R^2$  value of the ICER

expressed in QALYs was  $R^2=0.9654$ . The higher the  $R^2$  is the better our linear model fits for a given set of observations.

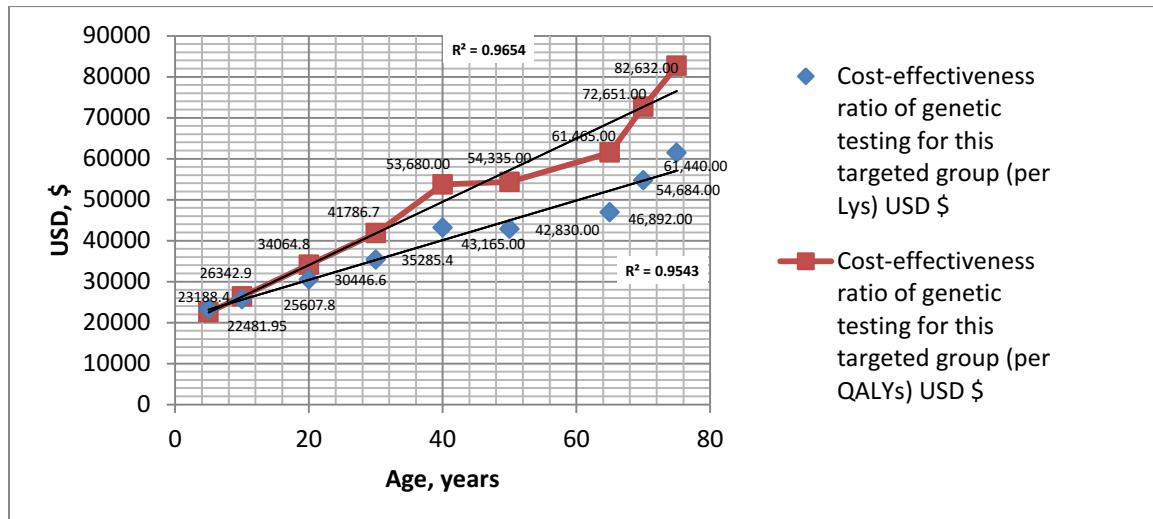


Figure 9. Aggregated outcomes of the ICER

### Model validation

To further confirm the linearity of the relationship in our model, I developed a plot model reflecting the rate of clinical encounters resulted from unintended drug-related injuries (Figure 10). Data points were gathered from previously published data (Budnitz et al., 2006) (Bourgeois, Shannon, Valim, & Mandl, 2010) (Bates et al., 1997). The authors collected these observations based on retrospective review of clinical encounters in emergency departments (Budnitz et al., 2006) and outpatient clinics (Bourgeois et al.,

2010) over several years in United States.

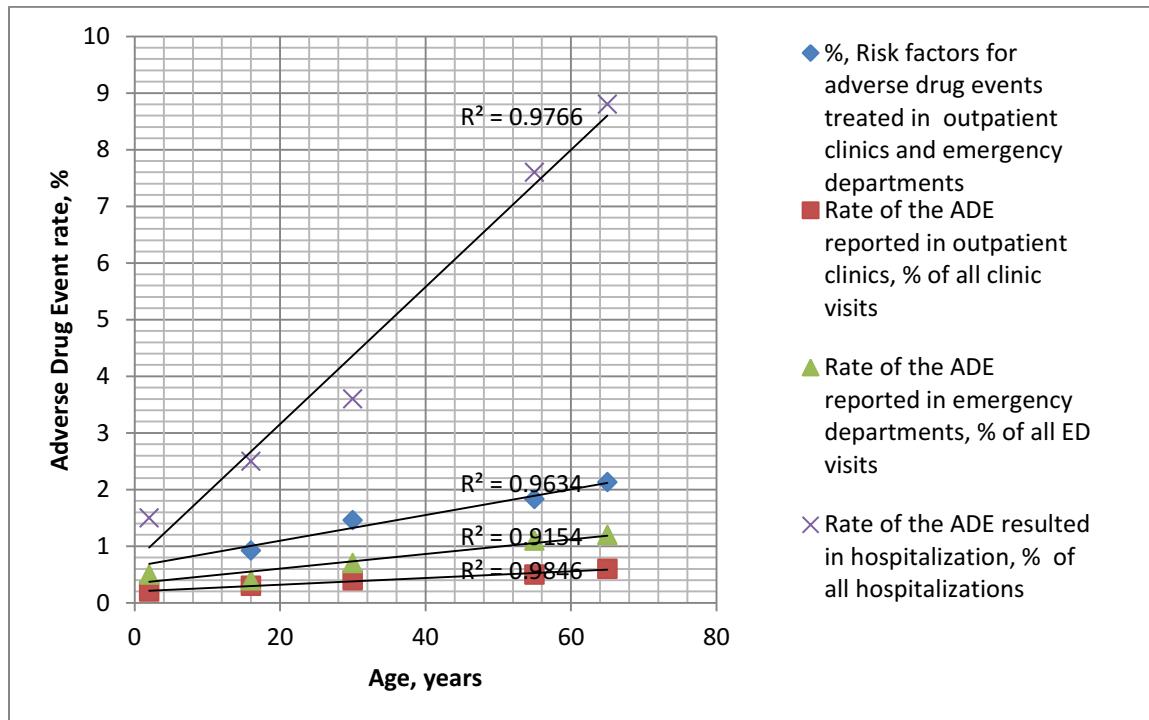
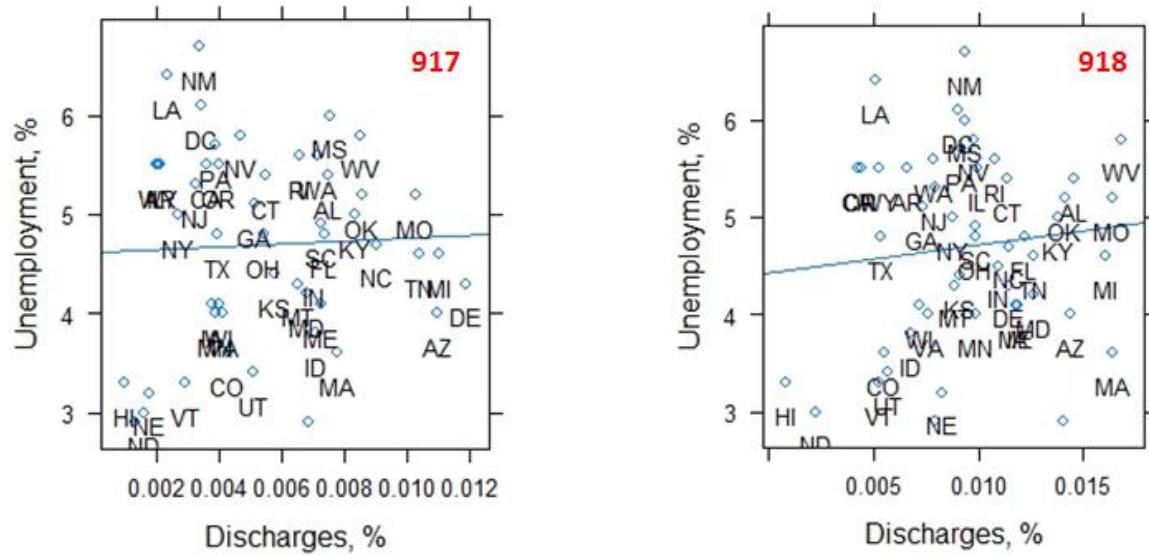


Figure 10. Estimated annual Incidence of Adverse Drug Events

Note: Values to develop this plot were adopted from (Budnitz et al., 2006) (Bourgeois et al., 2010) (Bates et al., 1997)

## Other contributing factors

To assess the possible correlation of clinically unrelated contributing factors, which might influence rates of ADE, I have designed a scatterplot correlation model of unemployment rate and household income to frequency of clinical encounters resulted from ADRs (Figures 11 and 12).



917: intercept 4.615; slope 15. 268

Correlation 0.04 (very weak)

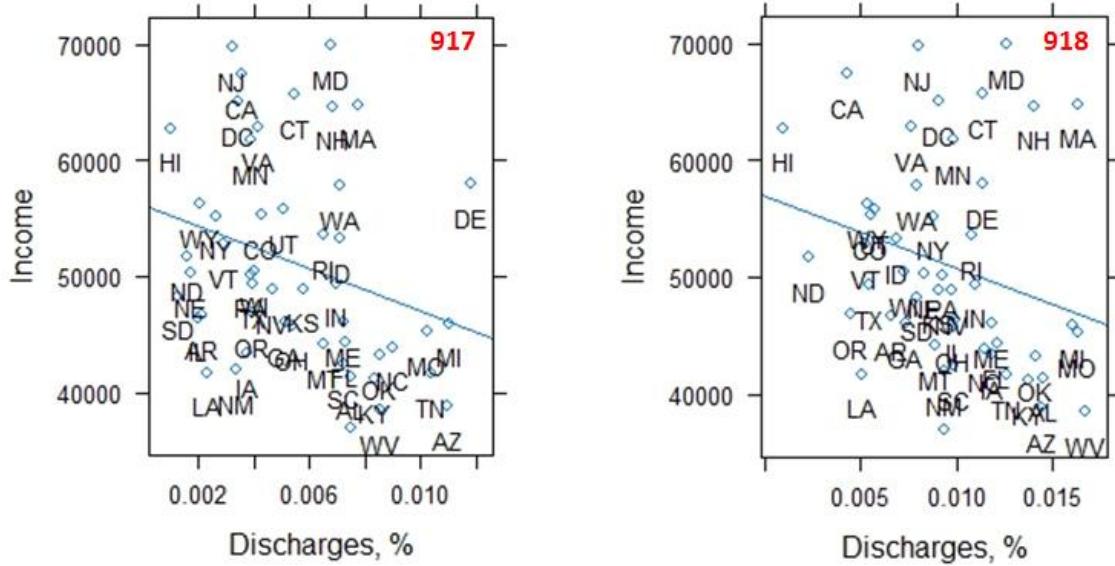
918 : intercept 4.42, slope 29. 16

Correlation 0.11 (very weak)

**Figure 11. Unemployment rate vs percentage of discharged with ADEs**  
Scatter plot: X-axis: percentage of people discharged with the condition of interest in each state; Y-axis: percentage of unemployment rate in each state, each dot represents given state.

In both models I detected a very weak but somewhat positive correlation between the unemployment rate in each state and the number of ADE-associated hospitalizations.

This correlation is not strong enough to suggest a conclusive relationship between the two variables.



917: Intercept 56156, Slope -915865

Correlation: -0.29 (moderate)

918 : Intercept 56975, Slope -616483

Correlation -0.26 (moderate)

Figure 12. Household income vs. clinical encounters due to the ADE  
Scatter plot: X-axis: percentage of discharges with condition of interest in each state; Y-axis: average income rate in each state, each dot represents given state.

In both models I detected a negative correlation between the percentage of discharges with condition of interest in each state and the average income rate. In the United States, hospitalizations due to ADEs are included in the top 50 most common diagnoses .Based on the initial calculations, costs associated with poisoning and toxic effects of drugs (with and without major complications) total to ~\$15.3million for ~46 thousand patient encounters (Figure 13).

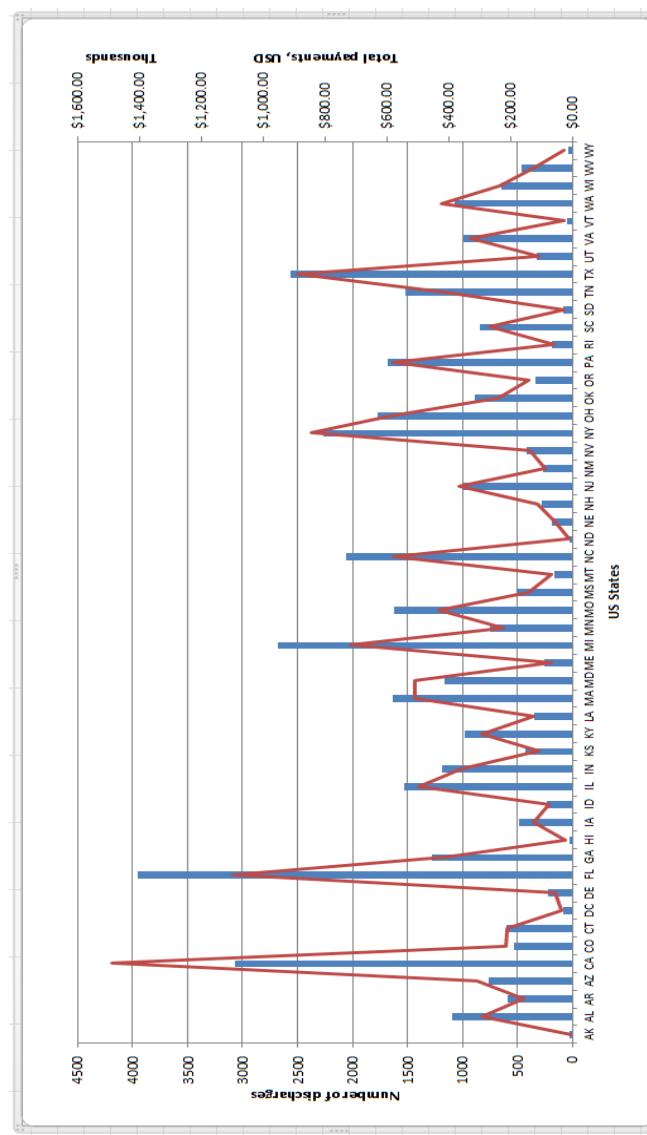


Figure 13. Frequency and cost associated with occurrence of ADEs

Expressed in number of ADE and total payments for clinical encounters (by US state).

This table visualizes financial burden and frequency of occurrence of ADEs.

Y-axis: (Left) number of discharged patients with ADEs (Blue bars)

Y-axis: (Right) Total payments done by payers for the (Red plot)

X-axis: US states

## Conclusion

A base line scatter plot to assess the relationship between ICER output data-points expressed in QALY and LY's was successfully developed (Alagoz et al., 2016) (Figure 8). The strength of the linear relationship on the graph was assessed based on the  $R^2$  values of QALY data points ( $R^2 = 0.8125$ ) and LY's data points ( $R^2 = 0.7643$ ). These  $R^2$  values were presumed to be close enough to the terminal value of 1 (or 100%) to be considered significant evidence for linearity of the plot figure.

Based on the outcome from the base line scattered plot that confirmed linear dynamics of the data set, I was able to extrapolate QALY and LY's values for the younger group of patients (Table 3). ICER expressed in QALY and LY's for the youngest group (5 years old) demonstrated significant decrease (\$ 23,188.04 per LY's and \$22,481.95 per QALY). Based on these values I conclude that we might achieve a 50% reduction in cost associated with preventable adverse drug events if we offered patients preemptive pharmacogenomic testing earlier in life.

To assess if the linearity of the relationship between the data points was preserved, I have visualized projected values on the scatter plot. Next,  $R^2$  was calculated for each series of data points to evaluate the strength of the regression line. The  $R^2$  value of ICER expressed in LY's was  $R^2=0.9543$  and the  $R^2$  value of the ICER expressed in QALYs was  $R^2=0.9654$ . These values indicated that our projected data points lay in very close proximity to the fitted regression line (Figure 9).

To further confirm the linearity of the relationship in our data points, a plot model reflecting the rate of clinical encounters attributed to unintended drug-related injuries was developed. The plot indicated a strong linear relationship between clinical encounters

associated with ADR and patient's age. The  $R^2$  values for all data series that represented rates of ARD-related clinical encounters at the emergency rooms, outpatient clinics and hospitals, ranged from 0.91 and 0.98 – all suggesting that plot has a linear dynamics of the data points.

Lastly, other unrelated, but potentially contributing factors were evaluated to determine possible correlation (Figure 12, 13). Neither unemployment rate nor household income demonstrated a sufficient correlation to with the rates of ADEs to suggest relationship.

## Chapter IV

### Discussion

Preemptive screening has been a reliable tool in the arsenal of healthcare providers for a long time (Birbeck, 2000). Clinicians and scientists univocally agree that validated and well-coordinated screening campaigns reduce morbidity, mortality and healthcare-related cost (Cohen, Neumann, & Weinstein, 2008).

Genomics and related modalities have proven to be trustworthy screening tools. Increasingly, genetic screening is part of standard of care and the first line of diagnostic approaches to explain disease-causing processes (Khromykh, A & Solomon BD, 2016; A. Khromykh & Solomon, 2015). In recent years, genetic and genomic testing has shown great potential on the preventive side of clinical care. Specifically, significant data exists to demonstrate pharmacogenomics' utility in preventive medicine (VanderVaart et al., 2011). PGx screening has not only demonstrated the ability to prevent adverse reactions to medications, it's been proven to guide therapeutic decisions on personalized dose adjustment of the pharmaceutical agents (Chang, Weitzel, & Schmidt, 2015). Moreover, research indicates that upfront investment in preemptive pharmacogenomic screening might decrease cost associated with healthcare services, relieve providers from an elevated number of patient encounters associated with adverse drug reactions and overall support the transformation of healthcare from reactive to predictive.

The goal of this work was to evaluate if it is more or less economically

advantageous to offer preemptive pharmacogenomic screening earlier in life. Statistical simulation models and pertinent calculations have indicated a strong connection between the timing of the screening and the possible savings of health-related costs. Data suggests that in some cases savings of up to 50% in prevented clinical encounters might be achieved if preemptive screening is offered in early childhood.

U.S. Food and Drug administration (FDA) (FDA.gov, 2016) reports that ADRs are accountable for \$136 billion yearly in health-related costs. According to O Alagoz et al, 17% of the ADRs could be prevented with preemptive pharmacogenomic screening if offered at the age of 40. That would result in savings of over \$23 Billion in health-related costs. The outcomes of this work suggest that the amount of savings could be over 38 billion dollars (12 Billion more) if such screening is offered at the beginning of one's life. Naturally, the amount of savings forecasted here is only an estimation of the actual amount of savings that will be achieved with preemptive PGx screening. Calculations described in this work were done based on the output data of the O Alagoz et al study developed for adults. The output data extrapolated to predict pediatric outcomes might not be sensitive enough to the potential variability in some segments of the considered variables. One of the future directions for this project could be development of predictive statistical models specifically for the early childhood group. Such a model could provide more accurate estimations of health-related savings and may reveal unexpected trends in data.

Another consideration for further research in the realm of pediatric PGx screening is an in-depth retrospective analysis of genotyping data if an ADE occurred at an early age. It has been reported in medical literature that pharmacokinetics of medications differ

widely between pediatric and adult populations. Such variability is in part a result of physiological differences, immaturity of enzyme systems, and clearance mechanisms in the child's liver (Ginsberg et al., 2002). Even with the insight of pre-emptive pharmacogenetic screening, healthcare providers have to be extra cautious when prescribing medications to young children due to their lack of fully functioning drug-metabolizing enzymes.

However, despite these warnings, preemptive pharmacogenomic screening can and should be offered during infancy. Similar to the blood type and newborn screening for metabolic diseases, PGx screening is destined to be a ubiquitous clinical tool. Early adoption of such screening should be an utmost priority of health centers around the United States and world-wide.

## References

- Alagoz, O., Durham, D., & Kasirajan, K. (2016). Cost-effectiveness of one-time genetic testing to minimize lifetime adverse drug reactions. *The Pharmacogenomics Journal*, 16(2), 129-136. doi:10.1038/tpj.2015.39 [doi]
- Alan Schwartz, A. E. (2001). MHPE 441: Medical decision making: cost-effectiveness analysis. Retrieved from <http://araw.mede.uic.edu/~alansz/courses/mhpe441/week10.html>
- Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12(5), 363-376. doi:10.1038/nrg2958 [doi]
- Amarasinghe, K. C., Li, J., & Halgamuge, S. K. (2013). CoNVEX: Copy number variation estimation in exome sequencing data using HMM. *BMC Bioinformatics*, 14 Suppl 2, S2-2105-14-S2-S2. Epub 2013 Jan 21. doi:10.1186/1471-2105-14-S2-S2 [doi]
- Azimi, N. A., & Welch, H. G. (1998). The effectiveness of cost-effectiveness analysis in containing costs. *Journal of General Internal Medicine*, 13(10), 664-669.
- Bates, D. W., Spell, N., Cullen, D. J., Burdick, E., Laird, N., Petersen, L. A., . . . Leape, L. L. (1997). The costs of adverse drug events in hospitalized patients. adverse drug events prevention study group. *Jama*, 277(4), 307-311.
- Birbeck, G. L. (2000). The benefits of screening must outweigh the risks and costs. *The Western Journal of Medicine*, 172(5), 308-309.
- Bodain, D. L., Solomon, B. D., Khromykh, A., Thach, D. C., Ramaswamy, K. I., Link, K., . . . Niederhuber, J. E. (2014). Diagnosis of an imprinted-gene syndrome by a novel bioinformatics analysis of whole-genoma sequences from a family trio. *Molecular Genetics & Genomic Medicine*, doi:Advance online publication. DOI:10.1002/mgg3.107
- Bodian, D. L., Klein, E., Iyer, R. K., Wong, W. S., Kothiyal, P., Stauffer, D., . . . Solomon, B. D. (2015). Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, doi:10.1038/gim.2015.111 [doi]
- Bodian, D. L., Solomon, B. D., Khromykh, A., Thach, D. C., Iyer, R. K., Link, K., . . . Niederhuber, J. E. (2014). Diagnosis of an imprinted-gene syndrome by a novel

- bioinformatics analysis of whole-genome sequences from a family trio. *Molecular Genetics & Genomic Medicine*, 2(6), 530-538. doi:10.1002/mgg3.107 [doi]
- Bourgeois, F. T., Shannon, M. W., Valim, C., & Mandl, K. D. (2010). Adverse drug events in the outpatient setting: An 11-year national analysis. *Pharmacoepidemiology and Drug Safety*, 19(9), 901-910. doi:10.1002/pds.1984 [doi]
- Brixner, D., Biltaji, E., Bress, A., Unni, S., Ye, X., Mamiya, T., . . . Biskupiak, J. (2016). The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. *Journal of Medical Economics*, 19(3), 213-228. doi:10.3111/13696998.2015.1110160 [doi]
- Budnitz, D. S., Pollock, D. A., Weidenbach, K. N., Mendelsohn, A. B., Schroeder, T. J., & Annest, J. L. (2006). National surveillance of emergency department visits for outpatient adverse drug events. *Jama*, 296(15), 1858-1866. doi:296/15/1858 [pii]
- Chang, K. L., Weitzel, K., & Schmidt, S. (2015). Pharmacogenetics: Using genetic information to guide drug therapy. *American Family Physician*, 92(7), 588-594. doi:d12149 [pii]
- Cohen, J. T., Neumann, P. J., & Weinstein, M. C. (2008). Does preventive care save money? health economics and the presidential candidates. *The New England Journal of Medicine*, 358(7), 661-663. doi:10.1056/NEJMp0708558 [doi]
- Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *The New England Journal of Medicine*, doi:10.1056/NEJMp1500523 [doi]
- data.gov. (2016). Hospital charge data. Retrieved from <https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzz3>
- Dunnenberger, H. M., Crews, K. R., Hoffman, J. M., Caudle, K. E., Broeckel, U., Howard, S. C., . . . Relling, M. V. (2015). Preemptive clinical pharmacogenetics implementation: Current programs in five US medical centers. *Annual Review of Pharmacology and Toxicology*, 55, 89-106. doi:10.1146/annurev-pharmtox-010814-124835 [doi]
- Elgin Community College. Least-squares regression. Retrieved from <https://faculty.elgin.edu/dkernler/statistics/ch04/4-2.html>

FDA.gov. (2016). Whole genome sequencing (WGS) program. Retrieved from  
<http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/>

Gallego, C. J., Bennette, C. S., Heagerty, P., Comstock, B., Horike-Pyne, M., Hisama, F., . . . Veenstra, D. L. (2014). Comparative effectiveness of next generation genomic sequencing for disease diagnosis: Design of a randomized controlled trial in patients with colorectal cancer/polyposis syndromes. *Contemporary Clinical Trials*, 39(1), 1-8. doi:10.1016/j.cct.2014.06.016 [doi]

Generics and Biosimilars Initiative (GaBI). (03/12/2010). What is the incremental cost-effectiveness ratio (ICER)? Retrieved from  
<http://www.gabionline.net/Generics/General/What-is-the-incremental-cost-effectiveness-ratio-ICER>

Ginsberg, G., Hattis, D., Sonawane, B., Russ, A., Banati, P., Kozlak, M., . . . Goble, R. (2002). Evaluation of child/adult pharmacokinetic differences from a database derived from the therapeutic drug literature. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 66(2), 185-200.

Hazrati, S. (2015). Congruence of parental genomic variation by country and self-report race and ethnicity in ongoing maternal child genomic studies. Retrieved from  
[http://www.abstracts2view.com/pas/view.php?nu=PAS15L1\\_1533.256](http://www.abstracts2view.com/pas/view.php?nu=PAS15L1_1533.256)

health.gov. (11/17/2016). Adverse drug events: Overview. Retrieved from  
<https://health.gov/hcq/ade.asp>

Hennekam, R. C., & Biesecker, L. G. (2012). Next-generation sequencing demands next-generation phenotyping. *Human Mutation*, 33(5), 884-886.  
doi:<http://dx.doi.org/10.1002/humu.22048>

Illumina. (2014). Retrieved from <http://www.illumina.com/>

Inova.org. (2016). MediMap<sup>TM</sup> PGx testing at inova. Retrieved from  
<https://www.inova.org/itm/medimap>

An introduction to next-generation sequencing technology . (2014). Retrieved from  
[http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf)

J. Scott Armstrong, Fred Collopy. (1993). Causal forces: Structuring knowledge for time-series extrapolation. *Journal of Forecasting*, 12(2), 103-115.  
doi:10.1002/for.3980120205

Johnson, J. A., Elsey, A. R., Clare-Salzler, M. J., Nessl, D., Conlon, M., & Nelson, D. R. (2013). Institutional profile: University of florida and shands hospital personalized medicine program: Clinical implementation of pharmacogenetics.

*Pharmacogenomics*, 14(7), 723-726. doi:10.2217/pgs.13.59 [doi]

Khromykh, A & Solomon BD. (2016). Managing the brave new world of genomics. Retrieved from <http://www.festivalofgenomicscalifornia.com/managing-the-brave-new-world-of-genomics/>

Khromykh, A., & Solomon, B. D. (2015). The benefits of whole-genome sequencing now and in the future. *Molecular Syndromology*, 6(3), 108-109. doi:10.1159/000438732 [doi]

Larry Copeland. (10/9/2014). Life expectancy in the USA hits a record high. Retrieved from <http://www.usatoday.com/story/news/nation/2014/10/08/us-life-expectancy-hits-record-high/16874039/>

M. Nanavaty, S. Kaura, M. Mwamburi, A. Gogate, J. Proach, A. Nyandege, Z.M. Khan. (2015). The use of incremental cost-effectiveness ratio thresholds in health technology assessment decisions. *Journal of Clinical Pathways*, 2017.

McCaffrey, N., Agar, M., Harlum, J., Karnon, J., Currow, D., & Eckermann, S. (2015). Better informing decision making with multiple outcomes cost-effectiveness analysis under uncertainty in cost-disutility space. *PloS One*, 10(3), e0115544. doi:10.1371/journal.pone.0115544 [doi]

Michael Sullivan, III, Joliet Junior College. (2012). *Statistics: Informed decisions using data*, 3rd edition (3rd ed.) doi:ISBN13: 9780321568021

national cancer institute. (2016). NCI dictionary of genetics terms. Retrieved from <https://www.cancer.gov/publications/dictionaries/genetics-dictionary?cdrid=776887>

National Institute for Health and Clinical Excellence. (2009). Measuring effectiveness and cost effectiveness: The QALY. Retrieved from <https://web.archive.org/web/20090809124835/http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenesstheqaly.jsp>

Neumann, P. J., Cohen, J. T., & Weinstein, M. C. (2014). Updating cost-effectiveness--the curious resilience of the \$50,000-per-QALY threshold. *The New England Journal of Medicine*, 371(9), 796-797. doi:10.1056/NEJMp1405158 [doi]

NextGxDx.Pharmacogenetic panels. Retrieved from <https://www.nextgxdx.com/apps/search/#/?q=pharmacogenetic>

- Oppong, R., Jowett, S., & Roberts, T. E. (2015). Economic evaluation alongside multinational studies: A systematic review of empirical studies. *PLoS One*, 10(6), e0131949. doi:10.1371/journal.pone.0131949 [doi]
- Owens, D. K. (1998). Interpretation of cost-effectiveness analyses. *Journal of General Internal Medicine*, 13(10), 716-717.
- Pavlopoulos, G. A., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., . . . Iliopoulos, I. (2013). Unraveling genomic variation from next generation sequencing data. *BioData Mining*, 6(1), 13-0381-6-13. doi:10.1186/1756-0381-6-13 [doi]
- Pharmainfo.net. (2016). Pharmaceutical information. Retrieved from <http://pharmainfo.net/>
- Pirmohamed, M. (2001). Pharmacogenetics and pharmacogenomics. *British Journal of Clinical Pharmacology*, 52(4), 345-347. doi:1498 [pii]
- Porter, M. E. (2010). What is value in health care? *The New England Journal of Medicine*, 363(26), 2477-2481. doi:10.1056/NEJMp1011024 [doi]
- PR Newswire. (02/08/2016). Inova women's hospital offers MediMap™ genomic testing to newborns as part of our standard package of care. Retrieved from <http://www.prnewswire.com/news-releases/inova-womens-hospital-offers-medimap-genomic-testing-to-newborns-as-part-of-our-standard-package-of-care-300216555.html>
- Prenetics. (2016). Medications: One size does not fit all. Retrieved from <http://www.prenetics.com/igenes-patients/>
- Price Ball, M. (2012). Seeking diversity (Especially Families). Retrieved from <http://blog.personalgenomes.org/2012/11/29/seeking-diversity/>
- Raczy, C., Petrovski, R., Saunders, C. T., Chorny, I., Kruglyak, S., Margulies, E. H., . . . Tanner, S. W. (2013). Isaac: Ultra-fast whole-genome secondary analysis on illumina sequencing platforms. *Bioinformatics (Oxford, England)*, 29(16), 2041-2043. doi:10.1093/bioinformatics/btt314 [doi]
- Reed, T. (2014). Genomics 2.0: Putting inova on the map hospital chain says new personalized medicine field has become part of its DNA. *Washington Business Journal*, , <http://www.bizjournals.com/washington/print-edition/2014/02/28/genomics-20-putting-inova-on-the-map.html>.
- Sheffield, L. J., & Phillimore, H. E. (2009). Clinical use of pharmacogenomic tests in 2009. *The Clinical Biochemist Reviews*, 30(2), 55-65.

Solomon, B. D. (2014). Obstacles and opportunities for the future of genomic medicine. *Molecular Genetics & Genomic Medicine*, 2(3), 205-209.  
doi:<http://dx.doi.org/10.1002/mgg3.78>

The European Bioinformatics Institute. (2016). What is next-generation DNA sequencing. Retrieved from <http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna->

Tracy Connell. (2016). Inova women's hospital offers MediMap™ genomic testing to newborns as part of our standard package of care.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics : TIG*, 30(9), 418-426.  
doi:S0168-9525(14)00112-7 [pii]

VanderVaart, S., Berger, H., Sistonen, J., Madadi, P., Matok, I., Gijsen, V. M., . . . Koren, G. (2011). CYP2D6 polymorphisms and codeine analgesia in postpartum pain management: A pilot study. *Therapeutic Drug Monitoring*, 33(4), 425-432.  
doi:10.1097/FTD.0b013e3182272b10 [doi]

William S. Weintraub, David J. Cohen. (2009). The limits of cost-effectiveness analysis. *Circulation: Cardiovascular Quality and Outcomes*, 2(1), 55-58.  
doi:<https://doi.org/10.1161/CIRCOUTCOMES.108.812321>

[www.whitehouse.gov](http://www.whitehouse.gov). (2015). FACT SHEET: President Obama's precision medicine initiative. Retrieved from <https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>

Yates, M., Starnes. Examining relationships: looking for patterns and departures from patterns. Retrieved from  
<http://www.henry.k12.ga.us/ugh/apstat/chapternotes/sec3.3.html>

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. *BMC Bioinformatics*, 14 Suppl 11, S1-2105-14-S11-S1. Epub 2013 Sep 13. doi:10.1186/1471-2105-14-S11-S1 [doi]