Fall 2009

Performance-Based Quality Specifications: The Link between Product Development and Clinical Outcomes

Steven Short

PERFORMANCE-BASED QUALITY SPECIFICATIONS: THE LINK BETWEEN PRODUCT DEVELOPMENT AND CLINICAL OUTCOMES

A Dissertation

Submitted to the Graduate School of Pharmaceutical Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

(Pharmaceutics)

By

Steven M. Short

December 2009

PERFORMANCE-BASED QUALITY SPECIFICATIONS: THE LINK BETWEEN PRODUCT DEVELOPMENT AND CLINICAL OUTCOMES

By

Steven M. Short

Approved October 23, 2009

Carl. A. Anderson, Ph.D. Associate Professor of Pharmaceutics Graduate School of Pharmaceutical Sciences (Committee Chair)	James K. Drennen, III, Ph.D. Associate Professor of Pharmaceutics Associate Dean for Graduate Programs and Research Mylan School of Pharmacy (Committee Member)
Robert P. Cogdill, Ph.D. Director of Research Development College of Engineering University of Nebraska, Lincoln Lincoln, NE (Committee Member)	Ira S. Buckner, Ph.D. Assistant Professor of Pharmaceutics Graduate School of Pharmaceutical Sciences (Committee Member)
Frank D'Amico, Ph.D. Professor of Statistics McAnulty Graduate School of Liberal Arts (Committee Member)	J. Douglas Bricker, Ph.D. Dean, Mylan School of Pharmacy and the Graduate School of Pharmaceutical Sciences

ABSTRACT

PERFORMANCE-BASED QUALITY SPECIFICATIONS: THE LINK BETWEEN PRODUCT DEVELOPMENT AND CLINICAL OUTCOMES

By

Steven M. Short

December 2009

Dissertation supervised by Carl A. Anderson, Ph.D.

The design of drug delivery systems and their corresponding dosing guidelines are critical product development functions supported by clinical pharmacokinetic (PK) and pharmacodynamic (PD) data. Largely, the importance of variance and covariance in product and patient attributes is poorly understood. The existence of PK/PD diversity among myriad patient sub-populations further complicates efforts to gauge the importance of product quality variation. Nevertheless, a platform capable of evaluating the effects of product and patient variability on clinical performance was constructed. This dissertation was predicated on requests to re-define pharmaceutical quality in terms of risk by relating clinical attributes to production characteristics.

To avoid *in vivo* studies, simulated experimental trials were conducted using the model drug, theophylline, for which data and models could be acquired from the

literature. Where comprehensive data were unavailable (e.g., production variability statistics), initial estimates were acquired via laboratory-scale experiments. Model asthmatic patients were generated using Monte Carlo simulation and published population distributions of various anothropometric measurements, disease rates, and lifestyle factors.

Mathematical constructs for *in vitro-in vivo* correlations provide a linkage between Quality by Design (QbD) product and process models, PK/PD models, and patient population statistics. The combined models formed the foundation for Monte Carlo risk assessments, which characterized the risk of inefficacy and toxicity for dosing of extended-release theophylline tablets. Sensitivity analyses revealed that patient compliance and content uniformity significantly influenced the probability of observing an adverse event.

The Monte Carlo risk assessment platform defined the link between the critical quality attributes (CQAs) and clinical performance (i.e., performance-based quality specifications (PBQS)). The PBQS were subsequently utilized to generate process independent design spaces conditioned on inefficacy and toxicity risk. These design spaces, which directly account for the conditional relationships between product quality and patient variability, can be transferred to a specific process via models that relate process critical control parameters to the CQAs. Process Analytical Technology, therefore, can be integrated into the QbD production environment to control the safety and efficacy of the final product. This work demonstrated that process and product knowledge can be used to estimate the risk that final product quality imparts to clinical performance.

DEDICATION

This dissertation is dedicated to my wife, Jessica, my parents, Chuck and Lynne, my extended family, and most importantly, God.

ACKNOWLEDGEMENTS

Graduate school is a long, tortuous test of an individual's fortitude, perseverance, and merit. Interestingly, no single person can successfully overcome its trials and tribulations without the extensive support system that flanks each and every student. I would like to take this opportunity to recognize those who helped me through this rewarding chapter of my life.

I am grateful to my academic advisor, Dr. Carl A. Anderson, for the opportunity to be a graduate student in his laboratory. Over the past several years, he has unwaveringly believed in my capabilities as a young scientist. His continued tutelage is inspiring, and has bestowed unto me numerous invaluable skills that will serve me well both professionally and personally. I am thankful for the time I spent at Duquesne University under his direction.

My experiences and accomplishments as a graduate student were enhanced by the leadership and wisdom of Dr. James K. Drennen, III. Although officially listed as a committee member, Dr. Drennen was truly a second advisor to me. As a man of many appointments, it would be easy for him to give only when convenient. Instead, Dr. Drennen selflessly devotes his time and energy to advance the lives and careers of all students, a characteristic of which I benefited from on countless occasions.

This work would not have been possible if it were not for the extraordinary visions of Dr. Robert P. Cogdill. I first knew Bob when he was a senior graduate student and was fortunate enough to benefit from his subsequent tenure as Industrial Research Coordinator here at Duquesne University. It was during these times I came to appreciate his remarkable drive and creativity. Although I often struggled to keep pace with his

innovate thoughts, his momentum helped propel me to this point. Bob is truly a gentleman, a scholar, and a friend, and I am thankful for his sustained mentorship.

I would like to thank my other committee members, Dr. Ira S. Buckner and Dr. Frank D'Amico. Their willingness to devote time and energy on my behalf speaks volumes about their character. I am grateful for their numerous insights that helped mold the work that this document details.

I am thankful for the Duquesne University faculty who provided valuable academic lessons, priceless research suggestions, and numerous opportunities to collaborate, all of which have impacted this work in one way or another. I am also appreciative for the wonderful staff who keeps the Graduate School of Pharmaceutical Sciences operating efficiently.

It would not have been possible to survive life as a graduate student without my fellow colleagues. The tenacious work ethic, constructive criticism network, and moral support system that can be found in our lab are second to none. I am appreciative for the frequent discussions we had which influenced my work. The myriad achievements and disappointments we shared will never be forgotten.

I would like to express my gratitude to Jonathan Gortat and Purdue University, Elizabeth-Hata International, Kaiser Optical Systems, Inc., and FOSS NIRSystems, Inc. for their generous donations of resources and equipment.

To all of the scholars and scientists who paved the way, I, and everyone else who will follow, are forever indebted. It is your intellect and courage to attempt the unthinkable that led us to this point.

I would not be here today if it were not for my parents. I mean not to reference the obvious gift of life, but rather the incredible acts of devotion and sacrifice they make on my behalf. The love and encouragement they provide me with is truly a blessing. I can only hope to display, and impart to others, the strength and discipline they demonstrate.

Words cannot express the gratitude I feel towards my wife for how instrumental she has been during my graduate career. To begin with, she supported my decision to apply to graduate school, knowing full well it would require significant sacrifice from both of us. She managed to deal with the long hours, the recurrent travel, and the all too frequent evenings and weekends that were spent on homework, experiments, speeches, manuscripts, and most recently, this document. She was my sounding board, always there to offer a fresh perspective on how to handle those difficult situations that accompany academia. It was her daily vigor that kept me motivated throughout this journey. I am blessed to have her by my side.

Finally, I would like to acknowledge God for the innumerable blessings I have been granted with. All of my good fortune can be attributed to his grace and mercy.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ADME	Absorption, Distribution, Metabolism, and Excretion
ANDA	Abbreviated New Drug Application
API	Active Pharmaceutical Ingredient
AUC	Area Under the Curve
AUMC	Area Under the Moment Curve
BCS	Biopharmaceutics Classification System
BMI	Body Mass Index
cAMP	cyclic 3’,5’-adenosine monophosphate
CDC	Centers for Disease Control and Prevention
CDER	Center for Drug Evaluation and Research
CDF	Cumulative Distribution Function
CFR	Code of Federal Regulations
cGMP	cyclic 3’,5’-guanosine monophosphate
CGMPs	Current Good Manufacturing Practices
CMC	Chemistry, Manufacturing, and Controls
COPD	Chronic Obstructive Pulmonary Disease
CQAs	Critical Quality Attributes
CYP	Cytochrome P450
DPBO	Defects per Billion Opportunities
DPMO	Defects per Million Opportunities
EPA	Environmental Protection Agency
FDA	Food and Drug Administration

FEV_1	Forced Expiratory Volume in one second (FEV_1)
FMO	Flavin-Containing Monooxygenases
FOM	Figure of Merit
GMP	Good Manufacturing Practice
ICH	International Conference on Harmonisation
IVIVC	*In Vitro-In Vivo* Correlation
LOD	Limit of Detection
LOQ	Limit of Quantification
MCC	Microcrystalline Cellulose
MCS	Monte Carlo Simulation
MEC	Minimum Effective Concentration
MSC	Multiplicative Scatter Correction
MTC	Minimum Toxic Concentration
NAS	Net Analyte Signal
NASA	National Aeronautics and Space Administration
NCHS	National Center for Health Statistics
NDA	New Drug Application
NHANES	National Health and Nutrition Examination Survey
NHLBI	National Heart, Lung, and Blood Institute
NIR	Near-Infrared
NIRS	Near-Infrared Spectroscopy
OR	Odds Ratio
PAT	Process Analytical Technology

PBQS	Performance-Based Quality Specifications
PCCPs	Process Critical Control Parameters
PD	Pharmacodynamic
PDF	Probability Density Function
PK	Pharmacokinetic
PLS	Partial Least-Squares
PRA	Probabilistic Risk Assessment
QbD	Quality by Design
QTPP	Quality Target Product Profile
RMSE	Root-Mean-Standard Error
RMSEC	Root-Mean-Standard Error of Calibration
RMSECV	Root-Mean-Standard Error of Cross Validation
RMSET	Root-Mean-Standard Error of Testing
RPM	Revolutions per Minute
RSD	Relative Standard Deviation
R&D	Research and Development
SNV	Standard Normal Variate
SR	Sustained-Release
S/N	Signal-to-Noise
SUPAC	Scale-Up and Postapproval Changes
US	United States
USP	United States Pharmacopeia
USP-NF	United States Pharmacopeia-National Formulary

Chapter 1: Introduction

1.1 Statement of the Problem

"Personal" computers that were once so massive they occupied more space than a small house have now been reduced to roughly the dimensions of a sheet of letter paper, yet they are orders of magnitude more powerful, not to mention less expensive, than their predecessors. Automobiles, which were once so costly they were a luxury to the most affluent of individuals, have since been mass-produced at reasonable prices with increased reliability and operational efficiency, revolutionizing the means by which practically all people commute. These are simple and obvious examples of the myriad unfathomable achievements that rest in the chronicles of history.

Numerous industrial sectors boast of extraordinary long-term progress regarding their innovative products and technological developments. The pharmaceutical industry, however, might not be able to do so. Granted, pioneering products such as osmotic tablets, drug eluting stents, and targeted gene delivery devices have all been developed near the turn of the century. Novel goods, however, are not the exclusive gauge of advancement. By and large, the pharmaceutical industry lags well behind other manufacturing sectors in terms of "quality." A study published in 2007 based on available benchmarks reported that pharmaceutical manufacturers operate on a level of approximately 35,000 defective units per 1,000,000 produced,[1] which is intolerable considering that other sectors have already achieved Six Sigma production (i.e., 3.4 defects for every 1,000,000 units); this is likely a function of the specifications more so

than the product. Moreover, disastrous clinical outcomes, perhaps best epitomized by the tragic Vioxx incident, further dilute the universal standard of quality associated with the pharmaceutical industry.

The Food and Drug Administration (FDA) has candidly acknowledged the lack of innovation that has impeded the growth of the pharmaceutical industry. Accepting partial responsibility for the current state, the FDA released a number of reports, initiatives, and guidances in the past decade to transform regulatory oversight and outwardly encourage reform, innovation, and low-risk, affordable medical products. The modified regulatory architecture is risk-based; the degree of regulatory scrutiny is commensurate to the risk a particular product imparts to the public. Companies that are able to adequately demonstrate the safety and efficacy of their products using risk- and science-based approaches will, therefore, be granted more regulatory flexibility.

In certain industries, faulty merchandise can result in mere inconvenience and frustration for the consumer. Defective (i.e., substandard quality) pharmaceutical products, however, may be ineffective or toxic, either of which can have grave consequences. The quality of medical products is typically assessed by a series of tests that characterize the products as acceptable or unacceptable. Acceptability is gauged via lower and/or upper specification limits, thresholds set with respect to a target value, which are to be derived based on product safety and efficacy. Product that is within the range defined by the specification(s) is deemed acceptable. Despite all of the collective efforts to ensure low-risk medical products, quality in the pharmaceutical industry is currently evaluated in such a way that the relationship between product specifications and clinical performance is implicit at best.

Numerous controllable and uncontrollable factors affect the safety and efficacy of pharmaceutical products. Pharmaceutical production involves the integration of several unit operations, each of which functions via a number of confounding parameters/variables that, if not controlled, have the potential to negatively influence product quality. Manufacturing requires a thorough understanding of these parameters/variables, and an appreciation for their interaction. Human variability also complicates safety and efficacy. Not all humans respond identically to the same drug administered at an equivalent dose; numerous physiologic and pathophysiologic variables affect clinical performance. Thus, these patient-specific factors also must be indentified and accounted for.

Recognizing the potential severity of defective pharmaceutical products, a lack of understanding is typically addressed by increasing the strictness of the specification; in other words, the range that defines product acceptability is reduced in an attempt to mitigate the likelihood of an adverse event. This, however, is often counterproductive. First, if the specifications are not explicitly derived in relation to the product's safety and efficacy, they may be too strict. Excessive thresholds carry the risk of not approving a product based on the misunderstanding the process and/or product. Contrary, overly stringent specifications may result in wasted resources allocated to improve the precision of the manufacturing process, which unnecessarily inflates the price of the final product. Worse yet, the adjustments may be too lenient, especially for certain high-risk subpopulations that were not identified or adequately evaluated during clinical testing. This may be caught by the company during post-marketing surveillance, or may be revealed through a high-profile incident, such as a death attributed to the drug. The

former would most likely mean costly re-evaluations of the dosing guidelines, the manufacturing process, and/or the product itself, which may even need to be temporarily pulled from the market. The latter, however, would almost inevitably result in immediate cessation of production and sale imposed by a regulatory consent decree, and potentially severe legal ramifications.

It is evident that inaccurate or unsubstantiated specifications contradict efforts to provide high-quality (i.e., low-risk), affordable medical products. The FDA, the pharmaceutical industry, and the general public unanimously desire safe and efficacious medical products. Given that quality has not been explicitly assessed in terms of clinical performance, the current perception of pharmaceutical quality may very well be inaccurate, in part, due to the way in which quality is defined and the lack of understanding regarding material and process variability. Inconsistent definitions and uninformative appraisals of pharmaceutical quality have initiated requests to standardize its definition in terms of risk to safety and efficacy. This work demonstrates how pharmaceutical process and product understanding can be used to supervise manufacturing based on the quantifiable risk that final product quality attributes impart to clinical performance.

1.2 Hypothesis and Objectives

This dissertation is based on the central hypothesis that pharmaceutical process and product understanding can be simultaneously utilized to model the risk that final product quality imparts to clinical performance.

Given the central hypothesis, the objectives of this dissertation were to:

1. Combine patient-specific factors with a pharmacokinetic model and an *in vitro-in vivo* correlation model to simulate drug plasma concentration profiles, which will be used in combination with a pharmacodynamic model to estimate patient risk in terms of inefficacy and toxicity.
2. Utilize Monte Carlo simulation and probabilistic risk assessment modeling to estimate the impact that changes in pharmaceutical manufacturing variability impart to patients.
3. Demonstrate how the risk simulation platform can be used to determine the conditional risk of product variation on clinical performance for a model solid oral dosage system.
4. Generate design spaces that are conditioned on quantitative estimations of inefficacy and toxicity risk.

Further, the methodologies developed in this work were used as the basis to propose a hypothetical scenario that couples Process Analytical Technology with Quality by Design such that production can be maintained in a low-risk state.

1.3 Literature Survey

1.3.1 The Pharmaceutical Quality Revolution

W. Edwards Deming, one of the forefathers of the quality revolution, once said that quality "… means a predictable degree of uniformity (variation) and dependability at low cost suited to the market."[2] Quality is unquestionably a cornerstone tenet that influences commerce, but what exactly is "quality" and more specifically, what does it mean to the pharmaceutical industry? The International Conference on Harmonisation (ICH), an organization that seeks to unify the regulatory bodies of Europe, Japan, and the United States and their respective pharmaceutical experts in an effort to register pharmaceuticals for human use, has posited several definitions during the past decade. The ICH Harmonised Tripartite Guideline Q6A defines quality as "the suitability of either a drug substance or drug product for its intended use. This term includes such attributes as the identity, strength, and purity."[3] The ICH Harmonised Tripartite Guideline Q9 offers an alternative definition: "the degree to which a set of inherent properties of a product, system, or process fulfills requirements."[4]

Although a standardized definition of "pharmaceutical quality" has yet to be acknowledged, its significance (or the lack thereof) has. The general public is ultimately the customer of pharmaceutical goods (and, therefore, the ones who dictate the level of acceptable quality). It is often assumed, however, that they are unable to adequately appraise these products. Therefore, regulatory authorities have absorbed this function so that patients need not be concerned about the medications they consume. The FDA has acknowledged that maintaining a scientific framework that ensures quality (low-risk),

innovative, pharmaceutical products is a primary public health objective. The Agency's ultimate mission (as it relates to the pharmaceutical industry) is to guarantee that the products available in the United States are consistently high in quality.[5]

It is well understood that "quality cannot be tested into products; it should be built-in or should be by design."[6] As a result, numerous regulatory initiatives, reports, and guidances have been introduced in the past several decades to recommend (both binding and non-binding) procedures and precautions to help ensure quality. In 1978, the FDA published *21 Code of Federal Regulations (CFR), Parts 210 and 211*, which outline the minimum current good manufacturing practice (CGMP) methods as well as the facilities and controls to be implemented for the manufacturing, processing, packaging, and/or holding of pharmaceutical products.[7] Segments of Parts 210 and 211 have since been revised. Although the CGMP regulations predate ICH Q6A, the underlying objective is to guarantee that the final products adhere to the standards of quality set forth by the ICH document. The regulations do not, however, define quality, nor do they adequately detail how to achieve quality during development and production.[7]

The CGMPs for the 21st Century initiative was launched by the FDA in August of 2002 to "modernize" the Agency's role in overseeing pharmaceutical manufacturing and product quality. In addition to underscoring its regulatory responsibility, the initiative implicitly stresses the obligatory role of all vested parties to supply quality medical products. To accomplish this, the existing CGMP programs were scrutinized to help assemble the regulatory architecture for overseeing manufacturing quality, which includes the concepts of risk management and quality systems. This clearly marked the beginnings of a risk-based inspection and enforcement program where regulatory

resources were employed in a manner commensurate with the threat to public safety. Eventually, the final report entitled *Pharmaceutical CGMPs for the 21st Century – A Risk-Based Approach* was released in 2004 to connect the CFR regulations and the Agency's existing views regarding quality systems.[5]

A formal guidance for quality systems, *The Quality Systems Approach to Pharmaceutical CGMP Regulations*, was released in 2006.[8] This document introduces the *quality system* as it relates to the pharmaceutical industry, which is a strategic business plan that formalizes corporate functions to attain product/service (quality) requirements, customer approval, and continuous improvement to ultimately achieve " … the public and private sectors' mutual goal of providing high-quality drug product to patients and prescribers." Designed appropriately, a quality system mitigates the risk of inferior (i.e., recalled, returned, salvaged, defective) products reaching the general public. Consequently, it has the potential to alleviate regulatory oversight considering that the robustness of the system is used as a criterion for determining the necessary extent of supervision.[5] The quality system is also advantageous in that it establishes the foundation for implementation of key developmental activities such as Quality by Design (QbD), continual improvement, and risk management.[8]

Backtracking in time, the FDA introduced several influential reports/guidances during the 2004 calendar year. The first of these was the Critical Path Initiative (via *Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*), which intends to expedite the development (i.e., time-to-market) of novel, safe, and effective medical products.[9] This is to be accomplished, in part, by taking a different approach to product development, which includes the use of innovative

techniques (e.g., computer models) to investigate the manufacturability, safety, and efficacy of candidate molecules and/or drug products.

The advent of comparability protocols is also integral to the objectives of the Critical Path Initiative. Comparability protocols were established via the *Comparability Protocols Protein Drug Products and Biological Products – Chemistry, Manufacturing, and Controls Information* draft guidance in September of 2003.[10] A *comparability protocol* "is a comprehensive, detailed, written plan that describes the specific tests and studies, analytical procedures, and acceptance criteria to be achieved to demonstrate the lack of adverse effect for a specified type of CMC [chemistry, manufacturing, and controls] change that may relate to the safety or effectiveness of a drug product."[5] Hence, a manufacturer can meet the criteria for a more lenient reporting category provided that they sufficiently demonstrate how the amendment will alter product quality (and, therefore, the risk to public health). This is accomplished by exhibiting a thorough understanding of the drug, its manufacturing methods and controls, and the effects that the projected change(s) will have on clinical performance. In certain scenarios, the change may actually be implemented prior to receiving regulatory approval, potentially decreasing the time-to-market.[10]

September of 2004 brought the release of two documents that are central to the pharmaceutical quality revolution, the *Pharmaceutical CGMPs for the 21st Century – A Risk-Based Approach*[5] and the Process Analytical Technology (PAT) guidance.[6] The CGMPs for the 21st century was previously addressed as it relates to the CFR regulations. This report is also significant in that it helped launch the QbD Initiative to encourage sound development of (quality) pharmaceutical products.[5] The QbD architecture

encompasses "designing and developing a product and associated manufacturing process that will be used during product development to ensure that the product consistently attains a predefined quality at the end of the manufacturing process."[8] Subsequently, the FDA advocated that "under the QbD paradigm, quality is built into the final product by understanding and controlling formulation and manufacturing variables: testing is used to confirm the quality of the product."[11] The ICH has also weighed in on QbD, describing it as "a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management."[12] Comprised of several factors including identification of critical quality attributes (CQAs), risk assessment, and continuous improvement,[13] the most renowned component of QbD is perhaps design space.[12]

The second 2004 document was the PAT guidance.[6] *Process Analytical Technology* is "… a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality." The PAT guidance details a scientific, risk-based regulatory framework that encourages innovative pharmaceutical development, manufacturing, and quality assurance through enhanced process understanding. The intent is to instill confidence in those implementing novel approaches that innovation, when conducted properly, does not beget additional regulatory scrutiny. In fact, manufacturing processes can be controlled and validated using the concepts discussed within the guidance.

Although not a definitive indicator, validating pharmaceutical production processes helps to assure quality. Given its importance, the FDA recently released a draft

guidance titled *Process Validation: General Principles and Practices*.[14] *Process validation* "… is defined as the collection and evaluation of data, from the process design stage throughout production, which establishes scientific evidence that a process is capable of consistently delivering quality products." This assumes that (1) quality has been integrated into the process (recall that it cannot be tested into a product via in-process and final-product assessments) and (2) each unit operation is under control such that the products adhere to all design characteristics and quality attributes, which include the established specifications derived from predetermined (tolerable) process average and dispersion estimates.[7] All processes, even those founded upon risk- and science-based design approaches, should incorporate in-process controls to assure product quality. When controlled appropriately, batch-to-batch, lot-to-lot, and unit-to-unit variability is such that the units sampled meet or exceed the validation criteria, giving the manufacturer, the regulatory authorities, (and in essence, the patient) confidence that the distributed products accurately reflect the label claims. Effective validation requires (1) an appreciation of the sources of variability, (2) detectability and sensitivity to variation, and (3) the capacity to "control the variation in a manner commensurate with the risk it represents to the process and product."[14] With liberty, the final statement also implies that a certain degree of variability is tolerable; acceptability should be proportionate to the risk it presents to patients. Validation also intends to confirm that the process is robust to factors that have the potential to stifle yield (e.g., raw material, manufacturing, environmental, and/or analytical method variability). Reductions in market supply clearly hurt the manufacturer, but more importantly, they impact the public well-being.

Continuous improvement is an underlying theme in each of the aforementioned documents and is critical to enhancing quality. Although the theory of continuous improvement is tacit within regulatory guidances, its premise is obvious. *Continuous improvement* involves the ongoing effort to enhance the efficiency and effectiveness of services or manufactured goods, and, therefore, broadly encompasses efforts that facilitate understanding, innovation, and the availability of affordable, quality drug therapies. Philosophies and initiatives such as Total Quality Management, Lean (Manufacturing), and Six Sigma are integral components of continuous improvement and should not be overlooked within the context of pharmaceutical quality (www.asq.org). The ultimate goal of continuous improvement is zero-defect production, where all units (drug products) conform to the utmost level of quality.

1.3.1.1 A Specification-Based World

ICH Q9 states that the "manufacturing and use of a drug product, including its components, necessarily entail some degree of risk."[4] It is understood that a certain level of variability is inherent to production. Moreover, the potential sources of variability are extensive, which increases the complexity of manufacturing. Variability, in turn, imparts risk to those who manufacture, regulate, prescribe, or consume medical products. It is therefore necessary to account for its influence. Variability (in essence, quality) is commonly evaluated via in-process and end-product tests (e.g., blend uniformity and content uniformity analyses, dissolution testing), which are centered on predetermined specifications. A *specification* is "a list of tests, references to analytical procedures, and appropriate acceptance criteria, which are numerical limits, ranges, or other criteria for the tests described."[3]

Despite the importance of specifications, there is little guidance available detailing how they are to be determined. *Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances Q6A*[3] and *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products Q6B*[15] are arguably the foremost documents on establishing specifications for the endorsement of non-registered new chemical and biological drug products, respectively; specifications for delivery systems that utilize well-documented active pharmaceutical ingredients (APIs) are often outlined in pharmacopoeias, such as the United States Pharmacopeia-National Formulary (USP-NF) monographs. Within ICH Q6A and Q6B guidelines, specifications are cast as an integral component in the overall control strategy of ensuring the consistency and quality of drug products; the control strategy includes (but is not limited to) validated manufacturing processes and test procedures, good manufacturing practices (GMPs), and raw-material, in-process, and stability testing. ICH Q6A states that the thorough understanding acquired during product characterization should be used to establish specifications. Guidances only imply that specifications should be derived with careful consideration of process critical control parameters (PCCPs) and product CQAs. *Critical quality attributes* are the physical, chemical, biological, or microbiological properties or characteristics that are known to affect product quality whereas *process critical control parameters* are the process variables that affect product CQAs. As such, PCCPs and CQAs should be identified during product development and, subsequently, should be controlled within tolerable operational and performance limits using data obtained during clinical and toxicological studies.

Acceptance criteria for new drug products are typically proposed, with accompanying rationalization, by the manufacturer and are approved by the regulatory agencies. They are implemented not to fully characterize a product, but to substantiate its quality. It is important to underscore, however, that testing via well-designed specifications is not the entirety of quality assurance. As stated in ICH Q6A, "the quality of drug substances and drug products is determined by their design, development, in-process controls, GMP controls, and process validation, and by specifications applied to them throughout development and manufacture."[3]

Needless to say, specifications serve as surrogates for quality within the current paradigm. All marketed drug products, via representative sampling, have, therefore, conformed to the established specifications, providing assurance to the patients that they are receiving safe and efficacious products. That is not to say, however, that the practitioners and patients themselves cannot impart additional risk by, for example, deviating from established dosing guidelines; dynamics such as this are likely to exacerbate risks imparted by manufacturing. Quality assessments that utilize numerical acceptance ranges dichotomize quality as either acceptable or unacceptable; only samples that fall within the upper and lower specification limits are deemed fit-for-use. At best, all marketed units of a given product are then, by default, categorized as low-risk or quality outputs. This message is rather misleading given that risk can be readily communicated as a continuous rather than a nominal metric. Moreover, even when all other factors are equitable, risk is often disparate within patient populations, which, although designed not to, can provide a false sense of security to patients.

Although not always the case, the formulae for performance measurements and acceptance tests often assume that the response is symmetric with respect to the target or average value. This suggests that both positive and negative deviations from the target carry the same intrinsic risk, which, in reality, may be incorrect. Take, for example, capability, which compares process performance to established specifications. A common metric for estimating capability is the process capability index (Cpk)

$$Cpk = \min\left[\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right] \tag{1.1}$$

where μ and σ are the mean and standard deviation of the process, which is assumed to be normally distributed, and *USL* and *LSL* are the upper and lower specification limits, respectively. Cpk can take on values of $(-\infty, \infty)$ and is negative only when the process mean is beyond the specifications. While certain capability metrics are meaningful only when the process is centered (e.g., Cp or process capability), Cpk reflects that processes are not always on target.

The process capability index, which assesses the specification range in relation to the breadth of the process, is taken as the minimum of the two estimates; the estimates are equal when the process is centered. Cpk can be enhanced by (1) reducing short-term variability, (2) altering the mean, or (3) relaxing the USL and LSL. Cpk and/or process sigma (via the Six Sigma program) estimates are often used to convey the level of repeatability or quality for a production process. Cpk and process sigma differ only by a factor of 3.0; hence, a Cpk of 2.0 is equivalent to a process sigma of 6.0.[16] Both metrics have statistical interpretation in that they communicate the likelihood of observing a unit outside of the specification limits (sometimes referred to as the defect rate), which is typically expressed as defects per million observations (DPMO) or defects per billion

observations (DPBO). For example, a Cpk of 2.0 (i.e., process sigma of 6.0) corresponds to 2.0 DPBO. It is important to note that a process sigma of 6.0 corresponds to 3.4 DPMO when the 1.5 σ shift is applied to account for drifts in the process mean.[1] The corresponding DPMO or DPBO for any Cpk can be determined from a *z* table for the standard normal distribution.

Capability, therefore, addresses the issue of how well the process is controlled with respect to the specifications by quantifying the measurements that are within the limits. Assuming that the specifications are established to mitigate risk (and that the more in control a process is, the less risk it poses), Cpk and risk are inversely related. Quality and Cpk, however, are positively correlated. Recognizing that the USL and LSL are typically set equidistant from the target value, Cpk does not take into consideration whether the deviation is positive or negative with respect to the mean, despite the fact that it can analyze non-centered processes. Given that risk is likely asymmetric with respect to the process mean (e.g., positive and negative deviations from the nominal API level may pose different risks), Cpk can be misleading if quality is interpreted, for instance, in terms of clinical performance. While this is irrelevant in certain industries, this is critical in regards to pharmaceutical products.

1.3.1.2 21st Century Pharmaceutical Production

ICH recently released the Harmonised Tripartite Guideline Q8(R1) to address the development of pharmaceutical products and their associated manufacturing methods.[12] One of the key objectives of Q8(R1) is to offer direction for effectively reporting process and product knowledge obtained through risk- and science-based design and development efforts to regulatory officials. Additionally, issues that often provide evidence to

inspectors and reviewers that the applicants have achieved an enhanced understanding are highlighted. Enhanced product and process knowledge combined with effective reporting methods mitigate risk to public safety, which can lead to reduced regulatory oversight.

ICH Q8(R1) also supports the QbD initiative. Not only does process and product design help to establish specifications and manufacturing controls, they aid in the construction of the design space(s). A *design space* is "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality."[12] The design space(s) is included in the application package and is therefore reviewable. Since it is subject to approval, only intentional deviations outside of the design space constitute a post-approval change; inadvertent departures are potential bases for product failure.

Appendix 1 of ICH Q8(R1) contrasts two very different approaches to pharmaceutical development; the minimal (sometimes referred to as the traditional)[17] and the QbD approach. The two methodologies are compared side by side in Table 1.1 based on six different categories: overall pharmaceutical development, manufacturing process, process controls, product specifications, control strategy, and lifecycle management.[12] At the one extreme, the traditional methodology underscores a rather archaic approach to pharmaceutical production. This approach embraces many of the practices that helped spur the pharmaceutical quality revolution. On the other hand, the QbD methodology employs the guiding principles of the quality revolution and serves as a step towards the desired state (i.e., risk- and science-based design, control strategies centered on quality, real-time release, risk-based regulatory architecture). While neither methodology

accurately summarizes the whole of pharmaceutical production, it is a safe assumption that the current state is more aptly outlined by the minimal approach.

Table 1.1 Comparison of the minimal and QbD methodologies for pharmaceutical development. Amended from ICH Q8(R1).

Aspect	Minimal … The Current State?	QbD…A Trend Towards the Desired State
Overall Pharmaceutical Development	• Empirically driven • Inefficient experimental design	• Mechanistically and scientifically driven • Multivariate experimental design • Incorporation of PAT and design space
Manufacturing Process	• Fixed • Concerned with reproducibility	• Flexible (within design space) • Concerned with robustness and a quality control strategy
Process Controls	• In-process testing only when necessary • Off-line analyses	• PAT system • Feed forward and feedback via process and control models
Product Specifications	• Principle method of control • Derived according to batch data	• Integrated component of quality control strategy • Derived according to desired product performance
Control Strategy	• In-process and final product testing	• Risk- and science-based quality control strategy • Real-time release
Lifecycle Management	• Reactive; action taken typically only when problems arise • Post-approval change	• Continuous improvement

As pharmaceutical production progresses towards the desired state, it is apparent that both PAT and QbD will have, at least for the foreseeable future, a central role in assuring product quality. Figure 1.1 illustrates the role of design space with regard to risk- and science-based pharmaceutical manufacturing. PAT is integral to designing,

analyzing, and controlling processes.[6] As a thorough understanding of the product and process is acquired through the PAT system, PCCPs and CQAs are identified. Research and development efforts are essential in gaining product and process knowledge. Additional experimentation, however, increases the manufacture's investment, which is why the direction and extent of research needs to be carefully balanced with the potential benefits in order to provide affordable, safe, and innovative medical products to the public. Experimental design, which is a key aspect of PAT, should be utilized to obtain the maximum amount of information in as few trials as possible. Experimental design undoubtedly extends operational ranges well past that which is implemented during controlled production, which enhances sensitivity to the critical parameters/attributes. These data are then used to construct the n-dimensional design space, which is comprised of inter-related PCCPs (e.g., compaction pressure) and material attributes that are known to affect product quality. The perimeter of the hyperspace is established according to a thorough understanding of the inputs and their relationship to a given performance metric(s). Efficient design and modeling not only reduce the ambiguity surrounding the boundaries, they increase the robustness of the design space, which minimizes the potential for upset (i.e., product of unacceptable quality) due to unanticipated variation. The critical-to-quality variables are then monitored (via sensors) and controlled (via process and control models) to be within the limits of the design space. This ultimately ensures the desired level of product quality. Ideally, the design space would be defined according to clinical risk.

Figure 1.1. The inter-relationship between PAT, design space, and pharmaceutical manufacturing. Adapted from R.C. Lyon, Process monitoring of pilot-scale pharmaceutical blends by near-infrared chemical imaging and spectrosocopy, Eastern Analytical Symposium (EAS), Somerset, NJ, 2006.

Given the overall complexity of pharmaceutical manufacturing, the number of material, process, and product attributes/parameters that potentially influence quality is sizeable. Experimental design and risk assessment can be utilized to not only help identify those factors that affect product CQAs, but to rank the importance of the individual factors based on their significance and elucidate potential confounding attributes/parameters. As development efforts progress and more data become available, they can be incorporated to gain additional understanding. These data also help establish the processing conditions under which quality product is reliably produced. The corresponding operational ranges of the PCCPs (as they relate to CQAs) can then be used to construct the design space.

As ICH Q8(R1) outlines, considerable regulatory flexibility is promised to sound submissions that provide justification for the inputs and process parameters that were included in (and excluded from) the design space.[12] A number of sequential design spaces can be generated for each unit operation within a manufacturing process. Figure 1.2 was reproduced from ICH Q8(R1), Appendix 2, Example 3.[12] It illustrates a design space for drying where the target moisture content is 1-2 %. Assuming that there were

additional unit operations downstream from drying, intermediate material outside the design space limits would not be suitable for further processing due to excessive impurity formation or excessive particle attrition. Alternatively, a single design space that bridges all unit operations is also acceptable. Figure 1.3, which was reproduced from ICH Q8(R1), Appendix 2, Example 2,[12] is an example of one such overarching design space that was constructed from two single attribute design spaces for friability and dissolution testing. The limits are simply defined as the overlapping ranges of tolerable outputs, which were both dependent on the two PCCPs. Given that both friability and dissolution are dependent on a number of inter-related factors, for example, the type and proportion of excipients, the homogeneity of the blend, the radial tensile strength of the tablets (which itself is dependent on a number of confounding manufacturing parameters, such as compaction pressure, turret speed, blending speed, blending time, charge order, blender volume, etc.), it is easy to conceptualize how this one design space accounts for numerous processing parameters and product attributes. Regardless of whether several consecutive hyperspaces or one overarching hyperspace is utilized, manufacturers can specify if they will operate within the entire design space or within some restricted area. Despite the flexibility, one thing should always be consistent from design space to design space: operation within the limits yields quality product.[12]

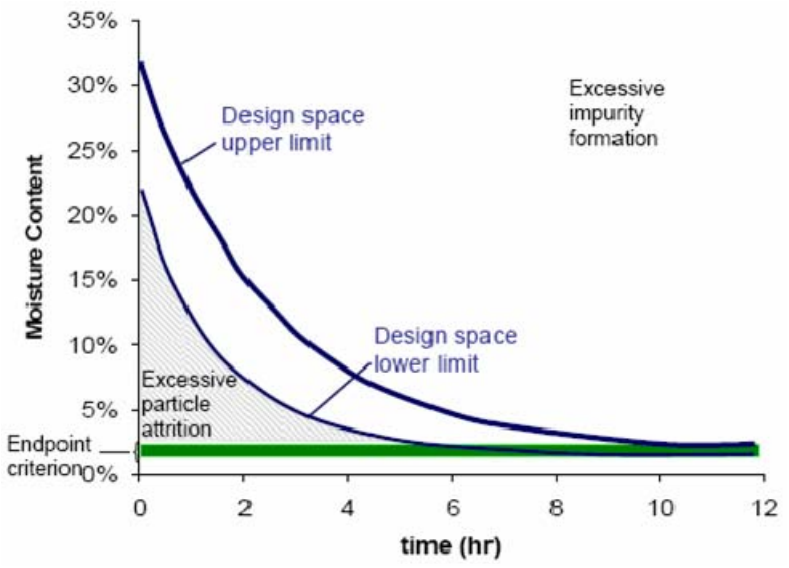

Figure 1.2. Example design space illustrating the moisture content as a function of time during drying. The target moisture content is 1 - 2 %. Intermediate material outside the limits would not be suitable for further processing. This figure was reproduced from ICH Harmonised Tripartite Guideline Pharmaceutical Development Q8(R1), November, 2008.

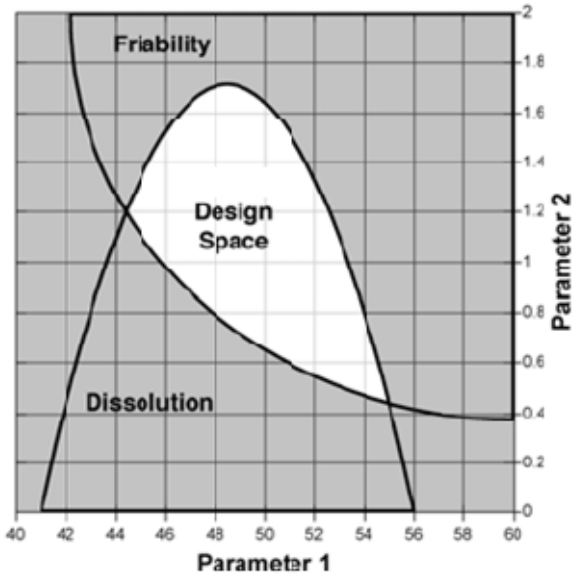

Figure 1.3. Overarching design space determined from two inter-related design spaces. Friability and dissolution were identified as the CQAs while two parameters (labeled 1 and 2) of a wet granulation unit operation were identified as the PCCPs (e.g., excipient particle size distribution, water content, granule size). The overlapping regions of the two inter-related hyperspaces are identified here as "Design Space." This figure was reproduced from the ICH Harmonised Tripartite Guideline Pharmaceutical Development Q8(R1), November, 2008.

Whether or not QbD concepts are integrated, the minimum compulsory components to be included in pharmaceutical development are:[12]

1. A quality target product profile
2. Potential product CQAs
3. CQAs of the API, excipients, and other integral components. This includes selection of the appropriate inactive ingredients and their proportions in the final delivery system.
4. A suitable manufacturing route
5. A control strategy.

These items are necessary to control the manufacturing process such that component and product CQAs are suitable to yield the desired quality target product profile (QTPP). A *quality target product profile* is "a prospective summary of the quality characteristics of a drug product that ideally will be achieved to ensure the desired quality, taking into account safety and efficacy of the drug product."[12] As such, product design and development efforts are structured around the QTPP. Overall, quality can not be assured when one or more of these elements are, in whole or part, absent.

While the above strategy can be followed to assure quality, the QbD initiative incorporates several supplementary strategies, which help lay the foundation for continuous improvement over the lifetime of the product. The QbD approach comprises:[12]

1. Systematically assessing and improving the product and process design, which includes

a. Use of *a priori* knowledge, experimental design, and risk assessment to identify material characteristics and process parameters that are potentially critical-to-quality
b. Determining, either empirically or via first principles, the mechanistic relationships between material characteristics and process parameters to product CQAs

2. Developing a control strategy founded upon process and product knowledge and quality risk management techniques. The control strategy may contain design space(s) and/or measures for real-time release testing.

Unlike final product testing, which as the name implies, involves assessing the product as or after it leaves the production line, *real-time release testing* is the assessment of in-process and/or final product quality via process data, which would consist of various PCCPs and CQAs. As quality is incorporated within the design, the product and its intermediates are assessed in real-time (i.e., during processing) to substantiate its quality. Real-time release therefore becomes a surrogate for the various analytical procedures necessary for final product release.[6]

The QTPP is clearly a crucial element in pharmaceutical development whether or not QbD tactics are incorporated. As consumers are more or less incapable of assessing quality, the manufacturer (by defining the QTPP) and the regulatory agency (by approving the product and its QTPP) ultimately dictate the level of quality that is acceptable for patients. The QTPP should be established according to numerous product and performance features, which include:

1. *"Intended use in clinical setting, route of administration, dosage form, delivery systems;*
2. *Dosage strength(s);*
3. *Container closure system;*
4. *Therapeutic moiety release or delivery and attributes affecting pharmacokinetic characteristics (e.g., dissolution, aerodynamic performance) appropriate to the drug product dosage form being developed;*
5. *Drug product quality criteria (e.g., sterility, purity, stability and drug release) appropriate for the intended marketed product.*"[12]

Since the QTPP specifies the desired quality and design space has a critical role in assuring product quality, it stands to reason that when pharmaceutical development involves QbD, the limits of the design space(s) can be defined by aspects of the QTPP. Moreover, the most important facet of the QTPP is the safety and efficacy of the drug product. Therefore, efforts should be made to construct the design space according to the attributes and parameters that directly affect safety and efficacy. After all, safety and efficacy are truly the principal benchmarks of pharmaceutical quality.

1.3.1.3 A New Definition of "Pharmaceutical Quality"

Quality is equally vital to the manufacturers, regulators, and consumers of pharmaceutical products. Regulatory officials, researchers, and academicians alike appreciate the disharmony surrounding pharmaceutical quality. In an effort to more effectively communicate allowable consumer risk, in part, determined through content uniformity testing, Williams *et al*.[18] suggested an alternative method where (risk) tolerance limits (for content uniformity) would be "based on a better understanding of

population and individual dose/response curves for efficacy and toxicity." Janet Woodcock, Director of the Center for Drug Evaluation and Research (CDER) at the FDA, pushed the envelope when she summarized several of the industry's (quality) shortcomings (e.g., quality assessment via end-product testing) and thus, the underlying frustrations of many individuals amid the "current state."[19] At the conclusion of her article, Dr. Woodcock purposed re-defining pharmaceutical quality in terms of risk by relating clinical attributes to production characteristics. The ideas put forth by Williams *et al*. and Woodcock are, at minimum, revolutionary as they not only present the opportunity to directly manage production in terms of product performance, they offer the impetus to reform the pharmaceutical industry. These visions have been largely ignored to date.

Cogdill and Drennen took notice when they laid the foundation for relating manufacturing quality and clinical performance of a drug product.[20] They described the combination of probabilistic risk assessment (PRA) and Monte Carlo simulation (MCS) to relate elements such as raw material quality, product design, population statistics, dosing guidelines, and patient compliance estimates with pharmacokinetic (PK), pharmacodynamic (PD), and *in vitro-in vivo* correlation (IVIVC) models to remold quality in terms of risk. The objective was to translate manufacturing and drug product attributes into estimates of risk of toxicity and inefficacy. Given the appropriate modeling, the interpretability of risk can be enhanced by transforming the metric from a nominal (i.e., acceptable or unacceptable) or categorical (i.e., low, medium, high) to a continuous (probabilistic) response. Moreover, the proposed platform departs from the current series of univariate measures that do not effectively account for covariance

among observations by harnessing multivariate data to simultaneously model the confounding factors that influence product performance. The model-based platform could then be used, for example, to define CQA hyperspaces bounded by risk or set manufacturing performance targets (e.g., Cpk) based on tolerable risk. With estimates of risk of toxicity and inefficacy, product and process design could then focus on minimizing overall risk to the patient. Risk to patients is both an effective gauge of quality and a more detailed means of communicating quality to consumers. This has the potential to redefine the marketplace, in effect reallocating purchasing power to the true customer (i.e., the patient), allowing him/her to make informed, risk-based decisions with the guidance of clinicians.

In addition to altering the manner in which consumers perceive pharmaceutical products, this novel methodology transforms the system in which products could be released and approved. The current system typically requires a series of sequential tests (e.g., content uniformity, assay, dissolution), which if the sampled units are within the specifications, the product is determined to be acceptable for release to the general public. Under the new paradigm, the confounding factors (e.g., uniformity, release rate, clinical performance) would be modeled (with respect to risk) simultaneously, effectively supplanting the need to set numerous specifications. Product release and approval, therefore, would be based on limits of tolerable risk to patient safety. This is consistent with the QbD approach of defining product specifications according the desired product (clinical) performance (Table 1.1).

Dickinson *et al*. published an article shortly after Cogdill and Drennen concerning dissolution testing and its link to clinical performance within the QbD paradigm.[21] The

objective was to demonstrate that *in vitro* dissolution testing is sensitive to manufacturing, formulation, and *in vivo* (bioavailability) changes such that the method could be used to assure that the safety and efficacy of newly tested products mirrors that observed in the preceding clinical assessments. A functional relationship between the pharmacokinetics and pharmacodynamics of a drug is truly required to effectively model clinical performance. As the authors highlighted, however, this relationship is not always available during the development stages. Thus, pharmacokinetic modeling was used as a surrogate for analyzing safety and efficacy. Ultimately, the relationship between manufacturing and composition variables and clinical performance was proposed to be used to set the boundaries of the operational design space for *in vitro* dissolution. The *in vitro* dissolution test, therefore, would be capable of assessing the influence that product changes have on clinical performance.

The authors proposed five strategic components for establishing a methodology capable of harnessing *in vitro* dissolution testing for evaluating clinical performance.[21]

1. Using *a priori* knowledge of the product under development and/or comparable products, perform a risk assessment to identify the various risks to clinical quality.
2. Develop a dissolution method that is sensitive to process and product variables that are expected to influence dissolution of the drug. The method should be physiologically relevant.
3. Ascertain from the *in vitro* and *in vivo* data the impact that changes to these variables have on clinical performance.
4. Define the dissolution specifications that ensure clinical performance.
5. Control dissolution within the limits of the design space to ensure clinical quality.

Following a risk assessment to prioritize the product and process variables most likely to affect safety and efficacy, four different dosage forms comprised of an unidentified model drug were investigated based on their likelihood to alter the dissolution release profile (which should affect the pharmacokinetics, and consequently, safety and efficacy). Following *in vitro* and *in vivo* analyses (in human volunteers), the authors observed comparable *in vivo* performances amongst the four formulations, despite their dissimilar dissolution profiles. Consequently, the authors did not attempt to establish an IVIVC, but concluded that it was not required to define a clinically relevant design space since *in vitro* dissolution testing was able to distinguish the formulation and manufacturing differences. Several different approaches to defining the dissolution specifications were presented. The first is in agreement with the FDA guidance[22] for situations where a Level A IVIVC has been established. When this is the case, C_{max} or AUC should not deviate more than ± 10 %. When the pharmacokinetics of the drug are not appreciably effected by variability in dissolution, as was the case for the model drug selected by Dickinson *et al*., the authors advocate that a "safe space" can be indentified where the change in pharmacokinetic parameters is minimal over a defined range of dissolution conditions (e.g., t_{50}, t_{90}). Lastly, a combination of a safe space and an IVIVC can be specified to allow the pharmacokinetic parameters to be controlled within ± 10 % deviation. Efforts such as these will reduce the burden of clinical trials, effectively promoting more efficient pharmaceutical development and manufacturing.[21]

1.3.2 Theophylline

1.3.2.1 Background

Theophylline (Figure 1.4) was formally introduced for the treatment of *status asthmaticus* (i.e., acute asthma) in 1937 following six years of investigational therapy during which Drs. Herrmann and Aynesworth observed relief in several patients.[23] It was, however, administered to humans as early as 1904, where it was noted to cause acute poisoning.[24] Since its introduction, theophylline has been used more frequently for chronic cardiorespiratory disorders[25] and, to a lesser extent, for the treatment of neonatal apnea.[26] A xanthine derivative, theophylline is prescribed for the management of acute and chronic bronchospasms linked to asthma and chronic obstructive pulmonary disease (COPD). Despite being largely supplanted by newer bronchodilators (e.g., corticosteroids, beta-2 adrenergic agonists), theophylline is beneficial for certain patients, including those who fail to adhere to or cannot meet demanding dosing schedules, and those who experience inadequate relief from available alternatives. It is one of the most commonly prescribed medications worldwide for the management of airway diseases due to its low cost. While it is generally regarded as a third-line treatment, theophylline is sometimes the only affordable alternative to inhaled corticosteroids.[27] Additionally, its anti-inflammatory and immunomodulatory effects have generated added attention to the drug's clinical potential.[25] Theophylline is also frequently utilized as a model drug for a variety of research foci, including investigations into the disparity in clearance mechanisms between neonates and adults[28] and the evaluation of physiologically-based PK models.[29,30]

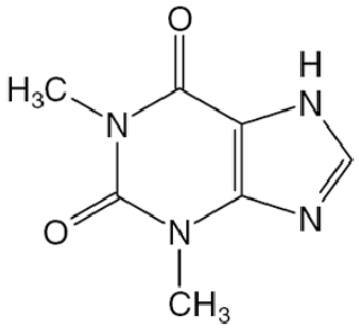

Figure 1.4. Chemical structure of anhydrous theophylline, $C_7H_8N_4O_2$, CAS registry number 58-55-9.

Theophylline has been almost universally classified as a bronchodilator, but its capacity to manage chronic asthma cannot be fully explained by its modest bronchodilator activity. Although the molecular mechanisms of action for theophylline are still in question, the general consensus is that the immunomodulatory, anti-inflammatory, and bronchoprotective effects collectively contribute to its pharmacologic actions. Theophylline nonselectively inhibits phosphodiesterase, which is responsible for cleaving phosphodiester bonds, such as those found in cyclic 3',5'-guanosine monophosphate (cGMP) and cyclic 3',5'-adenosine monophosphate (cAMP). At therapeutically relevant concentrations, theophylline induces an increase in intracellular cGMP and cAMP concentrations, which results in bronchodilation and inhibition of inflammatory and immune cells. Theophylline also acts as a nonselective antagonist of adenosine receptors at these concentrations. Although adenosine has virtually no effect on airway smooth muscle of non-asthmatic individuals, it elicits bronchoconstriction in asthmatic patients. Theophylline has also been noted to reduce fatigue in diaphragmatic muscles, enhance mucociliary clearance, block the decrease in ventilation that occurs with sustained hypoxia, and lessen microvascular leakage of plasma into the

airways.[25,27,31] Further information regarding the potential mechanisms of action is found in the articles by Barnes[27] and Barnes and Pauwels.[31]

1.3.2.2 Absorption, Distribution, Metabolism, and Excretion

Theophylline is a water-soluble drug that is rapidly and completely absorbed when administered in solution or as fast dissolving uncoated tablets.[25,28] Selected controlled- or extended-release dosage forms, however, are not 100 % bioavailable; certain products have reported fractions of the dose absorbed systemically of as low as 0.42.[32] Absorption occurs throughout the gastrointestinal tract, and the extent of absorption has been reported to be comparable in the stomach, the small intestine, and the colon.[33] Others, however, have observed greater quantities of theophylline absorbed in the small intestine when compared to the stomach and colon.[34] Theophylline is distributed primarily in extracellular fluid with a volume of distribution of approximately 0.45 to 0.5 L/kg. It is roughly 40 % protein bound (largely to albumin), but due to age-related variability, the range is reported to be 40 – 60 %.[35] Despite the variability, measurements of total serum concentrations are said to accurately reflect the unbound fraction.[36] Theophylline is typically considered to be rapidly removed from the body; the plasma half-life in adults is 7 – 9 hours.[35] While this may be accurate for the majority of the population, dramatic fluctuations in the elimination rate have been observed and the average half-life has been reported to be 1.2 – 65 hours.[37]

It is generally accepted that serum concentrations of 10 – 20 mg/L define the therapeutic range of theophylline.[25,38,39] Researchers, however, have observed responses at concentrations as low as 5 mg/L,[26,31] and thus, have suggested altering the therapeutic window to 5 – 15 mg/L to avoid potentially toxic events,[35] which are more probable at

higher serum concentrations. Others are content to report the therapeutic window as 5 – 20 mg/L.[40,41] Toxic effects of theophylline include nausea, vomiting, abdominal cramps, diarrhea, headache, arrhythmias, tremor, agitation, hypothermia, seizures, brain damage, and mortality.[25,42,43] Despite viewpoints that the severity of toxic events was correlated to the drug level, it is apparent that typical toxic symptoms associated with lower serum concentrations (e.g., nausea, vomiting) do not necessarily precede severe adverse reactions (e.g., seizure, death).[38]

Although theophylline has been used to manage cardiorespiratory conditions for over 70 years, the dose-response relationship remains unclear. Most references indicate that theophylline is primarily (~90 %) metabolized via the liver and the remainder is excreted unchanged in the urine.[35] While this is reasonably accurate for the general population, certain sub-groups (e.g., neonates, those with liver diseases) eliminate theophylline differently. Disposition, metabolism, and excretion of theophylline are dependent on a variety of factors and its metabolic pathway interacts with that of caffeine, which is also a xanthine derivative. Caffeine and theophylline are structural analogues that differ by a single N-methyl group and both molecules undergo hepatic metabolism via similar saturable mechanisms.[28]

The following discussion focuses on the metabolism and excretion of theophylline, and the relevant conversions between the two analogues. For particulars regarding the elimination of caffeine, please refer to the article by Ginsberg *et al*.[28] The predominant mechanism of removal for theophylline in adults is metabolism via the cytochrome P450 (CYP) 1A2 enzyme (Figure 1.5). The three resultant metabolites are 3-methylxanthine and 1-methyluric acid (via N-demethylation), and 1,3-dimethyluric acid

(via 8-hydroxylation); theophylline is converted to the intermediary product 1-methylxanthine, which is ultimately oxidized (via C_8-oxidation) to 1-methyluric acid by xanthine oxidase. 1,3-dimethyluric acid, 1-methyluric acid, and 3-methylxanthine constitute 45 – 55 %, 15 – 20 %, and 10 – 15 %, respectively, of the urinary metabolites in adults.[35] While metabolism is almost universally attributed to CYP1A2, other CPY enzymes (e.g, CYP1A1, CYP2E1, CYP3A4) and flavin-containing monooxygenases (FMO) have been established as minor contributors to the biotransformation of theophylline. For example, several studies indicated that CYP2E1 and CYP3A4, in addition to CYP1A2, contribute to the oxidative metabolism of theophylline to 1,3-dimethyluric acid.[44] Further, metabolism of theophylline by both CYP1A1 and CYP1A2 has been observed.[28,44] Given that only minor quantities of CYP1A1 are expressed in the liver, these findings suggest that a certain amount of theophylline is metabolized elsewhere. It is common modeling practice, however, to restrict metabolism to the liver only. Caffeine, on the other hand, counteracts the elimination of theophylline as one of its primary metabolic pathways yields theophylline (Figure 1.5).[28] Since the metabolism of caffeine to theophylline elevates blood levels of the latter, consumption of caffeine is often times restricted or even prohibited during treatment with theophylline. Consumption of soda, tea, and coffee has been reported to yield theophylline serum concentrations of less than 3 mg/L.[42]

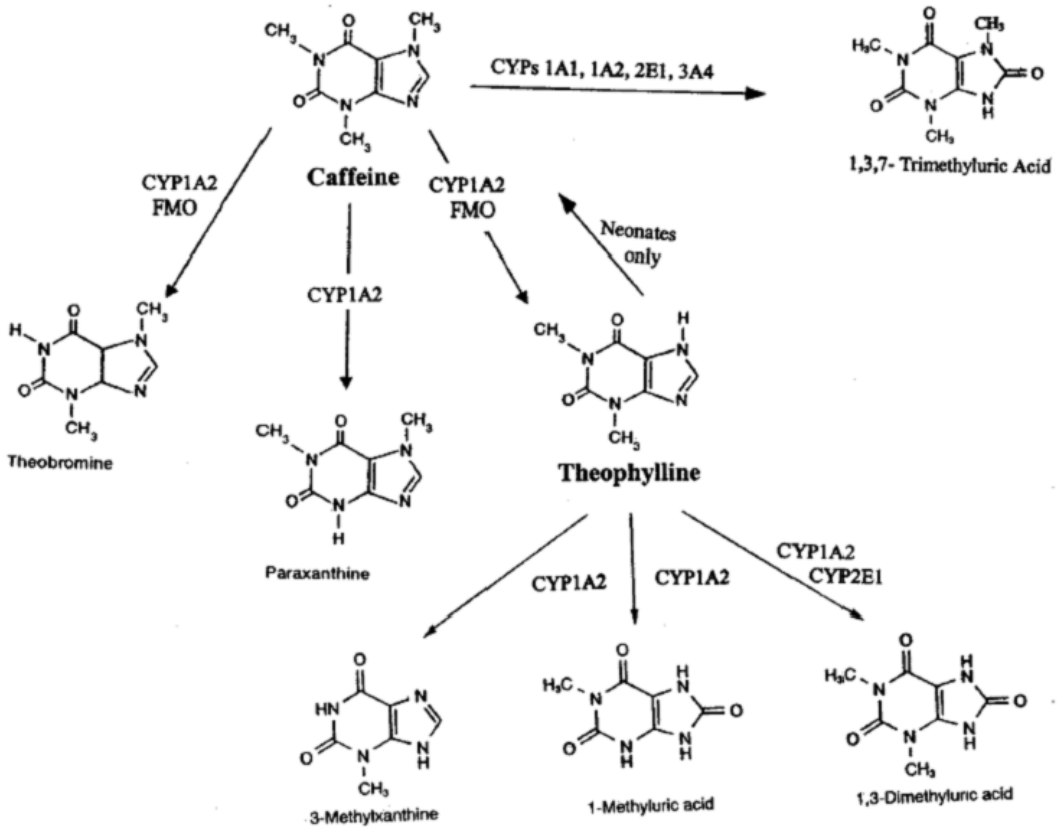

Figure 1.5. Diagram of the major metabolic pathways for theophylline and caffeine. CYP represents cytochrome P450 and FMO is the acronym for flavin-containing monooxygenases.

While predominant in adults, the CYP1A2 metabolic pathway is under-developed in neonates and infants. Consequently, theophylline elimination is altered in newborns. To begin with, roughly 50 % of the theophylline dose is excreted unchanged in the urine of neonates. Despite the fact that renal elimination is enhanced, it is not as efficient as metabolism via CYP1A2, which results in a net decrease in theophylline clearance in neonates as compared to adults. This is manifest by the three- to five-fold increase in the half-life of theophylline in neonates (20 – 34 hours compared to 7.3 hours in adults). Secondly, upwards of 40 % of the theophylline is metabolized via 8-hydroxylation in newborns (presumably by means of CYP2E1 and/or CYP3A4), which has been reported to be less than that in adults. Finally, approximately 5 – 10 % of the theophylline is

converted to caffeine and excreted in the urine, which has been observed in both neonates and infants.[35] This mechanism has never been detected in adults, suggesting that it is perhaps a compensatory pathway found only in neonates.[28,35] Metabolic enzyme activity comparable to that in adults is not normally observed until approximately 3 years of age.[44]

Inter-patient variability is perhaps best epitomized by the extreme disparity in theophylline clearance.[43,45] This is substantiated by the work of Jusko *et al*. who observed clearance values ranging from 4 – 143 mL/hr/kg in 200 individuals with diverse physiological, pathophysiological, environmental, and behavioral characteristics.[46] Factors other than age (e.g., sex, diet, concomitant drug therapies, recreational drug use, obesity, illness, pregnancy) also obfuscate the disposition and elimination of theophylline.[26,42,43,46,47] For example, polycyclic aromatic hydrocarbons found in tobacco smoke have been shown to induce several drug-metabolizing enzymes (i.e., CYP1A1, CYP1A2, and CYP2E1), which ultimately increase the clearance rates of theophylline by upwards of 80 % in smokers as compared to nonsmokers.[48] Elevated clearance rates have also been observed in smokers weeks to months following abstention. Individuals subjected to second-hand smoke also have enhanced theophylline clearances. In addition to the direct influence on enzymatic biotransformation, smoking has physiologic manifestations, such as vasoconstriction, which decreases blood perfusion to the liver, ultimately reducing the elimination of theophylline. Therefore, the safety and efficacy profiles for smokers are distinct, and their dosage regimens must be adjusted accordingly.[48]

1.3.2.3 Dosing

Numerous attempts have been made to define theophylline dosing regimens for given sub-populations,[37,41,42,49] but none have been proven to be applicable to the general population; when tested in independent sub-populations, most patients required dose adjustments to bring their serum levels within the defined therapeutic range.[43,45,47,49] Though Bayesian forecasting methods[50,51] appear to be most accurate, these dosing regimens do not supplant the need to closely monitor serum concentrations to ensure proper administration.[49] To date, no universal dosing method has been generated for theophylline.

Due to the unpredictable nature of dosage requirements, individualized dosage regimens have been posited as a reasonable method of safely dosing patients with theophylline.[35,39,49,52] "It is important to realize that there are wide interindividual differences in theophylline clearance and in the degree of its interaction with other agents. Therefore, data in the literature should be regarded only as a general guide, and careful observation for adverse drug reactions and blood level monitoring have to be conducted in most patients."[35] Individualized methods, however, are restrictive in that they necessitate estimations of serum concentration to ensure proper dosing. Hurley *et al*. demonstrated in a randomized clinical trial that in contrast to dosages based on serum theophylline concentrations alone, utilizing each patient's estimated theophylline clearance enhanced the accuracy of dosing.[53] Irrespective of the dosing scheme, extensive inter-patient pharmacokinetic variability coupled with the narrow therapeutic window complicate the administration of theophylline.

Several research groups have attempted to model the clearance of theophylline as a result of the significant challenges that inter-patient variability poses to clinicians for its accurate dosing. Chiou *et al*. introduced a method to rapidly estimate the total body clearance of theophylline following constant-rate intravenous (IV) infusion that requires two plasma samples and a reasonable estimate of the apparent volume of distribution.[37] This method was demonstrated using rabbits as test subjects. Similarly, Powell *et al*. assessed three routines for modeling total body clearance in healthy volunteers and patients suffering from airway obstruction who were administered theophylline via constant-rate IV infusion.[45] These methods required up to seven blood samples for the analysis of theophylline concentration.

In addition to modeling the clearance, Powell *et al*. also studied the effects of various (potentially confounding) factors on the clearance and volume of distribution, including sex, age, race, smoking habits, bronchitis, asthma, pneumonia, congestive heart failure, and severity of bronchial obstruction in 31 healthy volunteers and 26 patients. They found that sex, age, race, and the diagnosis of asthma or chronic bronchitis *per se* had no significant effect on clearance, whereas smoking, severe congestive heart failure, pneumonia, and severe bronchial obstruction resulted in significant changes. Additionally, the authors observed that, with the exception of patients who had severe congestive heart failure or pneumonia, changes in theophylline clearance during the course of the therapy were minor; no change as a function of time was observed in the healthy volunteers. In contract to clearance, the volume of distribution was not correlated with any of the factors examined. The authors suggested that these data supported the argument that changes in half-lives for a given disease state are a result of alterations in

clearance rather in the volume of distribution. Thus, with the exclusion of those patients whose clearance is expected to drastically change throughout the course of treatment, they conclude that a single theophylline concentration should be adequate for the entire IV theophylline therapy.[45]

Jusko *et al*. also performed an extensive retrospective study to model the environmental and pathophysiologic factors that affect theophylline clearance in 200 patients and normal volunteers, including age, sex, liver disease, congestive heart failure, obesity, renal function, history of drug, tobacco, marijuana, caffeine, or alcohol use, and pregnancy.[46] The history of drug use included consideration of oral contraceptives, barbiturates, benzodiazepines, phenothiazines, and tricyclic antidepressants. Clearance was modeled using a previously developed nonlinear algorithm.[54] The authors noted that other factors (e.g., marked dietary changes, respiratory viral illness, thyroid dysfunction, acute steroid administration, low arterial PO_2, recent ingestion of charcoaled foods) could affect the disposition of theophylline; however, they were either not acknowledged as factors at the time this study was performed or these data were unavailable. All but 8 of the subjects were Caucasian; therefore, differences attributable to race were not examined. Of the factors assessed, age, liver disease, smoking status, and congestive heart failure significantly altered clearance.

Additionally, Jusko *et al*. sought to characterize the variability in theophylline clearance within the sample population. Analysis of variance was used to determine the order, priority, and interactions of the factors that correlated with theophylline clearance to yield the maximum statistical discrimination between groups.[55] Clearance estimates were subdivided into mutually exclusive subgroups based on the reduction of

unexplained variance ($p < 0.01$). If a factor was previously accounted for, but had the potential to mediate the effects of factors yet to be considered, it could be reintroduced for assessment. Ultimately, the authors generated a clearance cascade, which illustrates the mean and standard deviation of total body theophylline clearance for each subgroup. This cascade can be reduced to a linear dosing nomogram to aid physicians in effectively targeting steady-state theophylline serum concentrations. The authors concluded that variability due to phenothiazines, tricyclic antidepressants, severe renal impairment, or oral contraceptives in nonsmokers could not be discounted as these categories were inadequately represented. Furthermore, they cautioned that many of the drug/disease/history/physiologic associations had not yet been verified in a prospective clinical trial.[46]

By means of 16 patients experiencing airway obstruction, Gilman *et al*.[56] compared the methods of Powel *et al*.[45] and Jusko *et al*.,[46] and a weighted least-squares Bayesian approach and the method of Chiou *et al*.[37] for the estimation of theophylline clearance following administration of aminophylline via IV infusion. Percent error was used to quantify the predictive error. Patients suffering from acute congestive heart failure, cirrhosis of the liver, pneumonia, sepsis, or severe malnutrition, or those who had taken erythromycin or cimetidine were excluded. The authors concluded that the Jusko *et al*. method out-performed that of Powel *et al*.; however, the patients were categorized into only 7 of the possible 16 terminal nodes of the Jusko *et al*. clearance cascade. Moreover, the Bayesian method was found to be more precise and less biased than the Chiou *et al*. routine; however, the requirement of selecting a prior model for the individualized expectations of pharmacokinetic parameters and their corresponding

variances limits the applicability of the Bayesian method when these models have not been adequately developed.

The most accepted dosing schedules for theophylline require administration one (i.e., q.d.), two (i.e., b.i.d.), or three (i.e., t.i.d.) times per day; however, regimens that necessitate multiple daily dosages are believed to exacerbate patient non-compliance. However, previous reports have shown that once-daily dosing of theophylline increases the likelihood of observing serum concentrations outside the therapeutic window.[32,57-59] This finding has lead researchers to advise twice-daily dosing, which has been shown to result in less extensive peak/trough fluctuations, thus increasing the probability of maintaining concentrations within the desired range. Despite the (supposed) reduced risk of an inefficacious or toxic event associated with twice-daily dosing, researchers contend that higher clinical efficacy is attained with once-daily dosing.[58,60,61] Additionally, it has been stressed that fluctuations in clinical effects that closely parallel the oscillations in serum concentration, which would advocate twice-daily dosing, have not been substantiated in published studies.[58]

1.3.3 Mathematical Modeling and Numerical Simulation

The terminology *modeling* (or model) and *simulation* are occasionally used interchangeably so as to suggest they share a common definition. While they both are abstract representations of real-world systems, each has a distinct meaning and it is important to differentiate between them. These terms are used throughout the remainder of this dissertation according to the definitions provided by Peter Bonate, noting that the original definition of simulation is expanded.[62] A *model* signifies any mathematical construct generated using fundamental processes or data that relates inputs to outputs,

whereas *simulation* is the exploitation of models to examine the long-term impact that variability [or uncertainty], treated as an input to the model, has on a system. The term simulation is broadened to include assessments of uncertainty. Table 1.2, which was adapted from Bonate's work, contrasts the two expressions.[62]

Table 1.2. Comparison and contrast of the terms modeling and simulation. Adapted from Bonate, PL. 2000. Clinical trial simulation in drug development. Pharmaceutical Research 17(3):252 – 256.

Modeling	**Simulation**
Sensitive to assumptions	Sensitive to assumptions
Sensitive to black-box criticisms	Sensitive to black-box criticisms
Uses data	Builds upon models based on data
Useful method for data summarization	Useful method to summarize complex inter-relationships between variables
Relates inputs to outputs	Incorporates random variability [or uncertainty] into a model and assesses its effects long-term
Random variability is a nuisance variable	Random variability can be incorporated in the simulation
Looks back in time	Looks forward in time
Can identify which variables are more important than others	Can indentify which variables are more important than others
Cannot be replicated	Can be replicated

1.3.3.1 Monte Carlo Simulation

Mathematical models provide users the opportunity to manipulate inputs to examine outputs, which offers greater insight into the functionality of processes and systems. Models are categorized according to the flexibility of their inputs; when all are fixed, the model is said to be *deterministic*, whereas when some or all of the terms are characterized by a certain level of random variability, the model is *stochastic*. Simulation harnesses models to observe an outcome or a prediction based on a given set of parameters (i.e., inputs). Stochastic modeling techniques that utilize random variability are typically referred to as MCS. Stanislaw Ulam and John von Neumann are credited

with coining the expression Monte Carlo, which came during their work on the atomic bomb at Los Alamos in 1946. Rugen and Callahan make the distinction, however, that Ulam developed the approach earlier while playing solitaire.[63] Though the phrase is a reference to the gambling casinos in Monaco, it was initially used to simulate random neutron diffusion in fissile material.[63-65]

The Monte Carlo method is a numerical technique used to solve mathematical functions by random sampling. Although the theory of random sampling was established long before the Monte Carlo method, MCS was not practical until the advent of computers. The universal approach of the Monte Carlo method is simple; MCS iteratively samples a set of random model parameters from their underlying distributions (which are generated *a priori*), performs a number of deterministic computations using the inputs, and stores the resultant outputs. Thus, a program is constructed to conduct a number of independent, random trials, the results of which are accumulated at the conclusion of each iteration. The Monte Carlo method is applicable to any system that is affected by random factors. Unlike deterministic functions, MCS has the distinct advantage of estimating the sensitivity or robustness of a system to random variation or error through the propagation of uncertainty. This uncertainty allows stochastic elements of a system to better represent practical observations, which can lead to more precise conclusions and/or actions.[64,65]

The Monte Carlo method utilizes random variables, which implies that for any given trial, the value assumed for a given input is unknown. The term "random," however, has a more specific meaning in the context of MCS in that the range of potential values the variable can assume, along with the probabilities of these values are

known *a priori*. Truly random numbers or distributions are cumbersome since it is often desirable to reproduce calculations performed on a computer; in order to repeat trials using random numbers, the numbers would need to be stored during each trial. Thus, MCS is typically conducted with what are known as pseudo-random numbers, which are values generated to approximate a particular distribution (e.g., uniform, normal) using mathematical formulae that simulate random numbers. Rather than having to test the validity (or randomness) of the random numbers at the start of each trial, pseudo-random number generators need only be tested once to confirm their ability to generate a sequence of numbers that approximates the properties of random values. The pseudo-random numbers can then be re-generated if need be using the same "seed" or point at which the algorithm began generating values.[64,66]

Pseudo-random number generators do have one significant limitation; the total number of values that can be generated using any given seed is finite. Eventually, identical values will be re-generated as the pseudo-random number generator cycles. Therefore, the user must verify that the sequence of pseudo-random values generated is large enough for the specific application such that earlier trials do not become correlated with later trials as a result of re-sampling. This ensures that the simulation is truly stochastic. Pseudo-random number generators are restricted by the number of bits (n) the computer possesses such that a sequence, often times referred to as the period, of pseudo-random numbers can be no longer than 2^n. Newer algorithms, however, have circumvented this issue, utilizing numerous generators for the seed and the integer generator such that periods of 2^{64} (Marsaglia's Ziggurat algorithm) and $2^{19,937-1}$ (Mersenne Twister algorithm) are attainable.[64,66]

The results of any MCS trial are dependent upon the number of times the system is perturbed. The model(s) must be tested a sufficient number of times such that the parameters sampled are representative of the general population. If the sampling is inadequate, the output will be misleading since certain input conditions that are likely to occur will not have been considered. While it is not possible to offer a general rule of thumb as to the required number of iterations, the simulation should iterate enough times to allow the output to stabilize. Ensuring that the change in the estimate drops below a defined threshold for a set number of consecutive iterations is one approach to verifying stabilization.[64,65]

1.3.3.2 Pharmaceutical Applications of Monte Carlo Simulation

Since its formal introduction, the number of applications that harness MCS has grown substantially. Monte Carlo simulation is utilized in many scenarios, including the general fields of manufacturing, economics, and science.[63] While a thorough review of all the applications of MCS is well outside the focus of this dissertation, the following discussion offers a brief overview of its major uses within the pharmaceutical industry. Additional detail regarding MCS is provided in section 1.3.7.

Monte Carlo simulation is a versatile tool that supports numerous aspects of the drug discovery and development process. The preliminary tasks of generating models and defining simulation inputs *a priori* force companies to identify the components that are well understood and, likewise, the elements that are uncertain or missing altogether. Although MCS is used throughout the drug discovery and development process,[67] its greatest impact as it relates to pharmaceutics is perhaps in the area of clinical trials, commonly referred to as clinical trial simulation.[67,68] To start with, MCS can be used to

study the economic investments necessary to conduct these trials, with the obvious objective of minimizing the total cost.[62] Poland and Wada utilized PK, PD, and economic models along with MCS to study the effects of the dosing regimen (i.e., once- versus twice-daily administration) and patient compliance on the efficacy of an antiretroviral drug in Phase II trials.[69] The authors concluded that despite enhanced patient adherence to the once-a-day dosing schedule, the marginal increase in long-term efficacy did not warrant the change from twice- to once-daily administration in view of the delayed time-to-market, additional monetary investments, and technical uncertainties associated with the new product. This is an excellent example of how MCS helped circumvent additional clinical trials that would have undoubtedly increased development expenses.

Many other facets of clinical trials can benefit from MCS. Although their explicit objectives may not have included minimizing financial investments, there is a consistent economic implication in the following cases. For example, MCS is used to approximate the initial dose in humans for phase I studies. This can be of particular benefit in allometric scaling efforts where PK parameters have been well-characterized (typically as averages) within the confines of animals, but transferability to humans is uncertain. Simulation can help elucidate which parameters have the largest effect on scaling the dose to humans by propagating variability through the models. Moreover, MCS is beneficial for determining the dose for phase II and III studies where the test subjects are typically patients rather than healthy volunteers. Therefore, fewer adverse events are anticipated from optimizing the dose *in silico* rather than *in vivo*, which inevitably increases the likelihood that the particular drug will receive approval for marketing.[62]

With a predictive PK model, single- and multiple-dose plasma versus time profiles can be generated with the aid of MCS, which can then be used to optimize the dose and the dosing schedule to ensure plasma concentrations are within the desired therapeutic range. Gomeni *et al*. took a similar (but extended) approach where they used PK and PD modeling in conjunction with MCS to assess the effects of PK absorption and disposition parameters (e.g., fraction of dose available), and inter-individual variability on percent receptor occupancy, which had been identified as a surrogate for efficacy.[70] The authors were able to identify the variables that had the largest influence on receptor occupancy, which enhanced their company's understanding of the unidentified drug and helped direct future development efforts.

As evidenced by the work of Poland and Wada, and Gomeni *et al*., models representing various components fundamental to clinical trials (e.g., PK and PD) can be integrated with MCS to investigate the effects of various inputs (e.g., patient compliance, inter-individual variability) on the clinical effectiveness of a given treatment. This offers the user the unique opportunity to explore numerous circumstances, some of which are not ethical in a clinical trial setting. For instance, MCS can be used to investigate the therapeutic outcome of patients missing a given percentage of their doses, which is not acceptable in situations where the patient would potentially experience significant discomfort or harm. The ability of simulation to efficiently address the "What if" questions *in silico* rather than *in vivo* underscores its utility.

Dutta and Reed utilized MCS to investigate the effects of patient compliance on plasma concentrations of valproic acid during treatment with 12-hour enteric-coated divalproex sodium tablets.[71] PK simulation is an ideal platform to assess the effects of

missed doses, and patient/clinician compensations for the omissions, seeing as how withholding medication from patients suffering from epilepsy or acute mania is unethical. Valproic acid, one of the two APIs in divalproex sodium tablets, has a narrow therapeutic window of plasma concentrations of 50 – 100 mg/L. Given its narrow therapeutic range and short elimination half-life of 6 – 16 hours, patient compliance is critical to the safety and effectiveness of this modality; low plasma concentrations can result in seizures whereas high plasma levels run the risk of inducing clinical toxicity.

The authors wanted to examine the PK effects of missing one or two doses in conjunction with three different temporary dosing regimen adjustments: (1) one missed divalproex dose compensated for by doubling the dose at the next scheduled dose administration (12-hour replacement), (2) two consecutive missed divalproex doses offset by administering a doubled dose 6 hours following the first missed dose (18-hour replacement), and (3) two consecutive missed divalproex doses counteracted by tripling the dose at the next scheduled dose administration (24-hour replacement). Moreover, they were interested in the confounding effects (i.e., drug-drug interactions) of hepatic enzyme-inducing antiepileptic drugs on valproic acid plasma concentrations. To accomplish this, adjustments were made for certain individuals to simulate the shortened elimination half-lives and elevated clearances observed for patients taking antiepileptic medications.

Population mean PK parameters for dose, bioavailability, absorption lag-time, first-order absorption rate constant, steady-state volume of distribution of unbound drug, protein binding parameters (i.e., the number of binding sites for the two classes of binding sites and their corresponding binding association constants), elimination half-life,

and clearance of free drug were taken from the literature. Different doses, elimination half-lives, and clearances were implemented for patients with enzyme induction versus those uninduced. The age range of the hypothetical patients was restricted to 11 – 64 years and body weight was assumed to be 70 kg. Patients not taking other enzyme-inducing antiepileptic drugs were administered 562.5 mg every 12 hours (i.e., 16 mg/kg/day), whereas induced subjects were administered 1125 mg (i.e., 32 mg/kg/day). Log-normal distributions representing 20 % inter-patient variability were generated for clearance, volume of distribution, albumin concentration, and the protein binding parameters using the population averages. Patients, in effect, were generated by randomly sampling the underlying distributions for the parameters estimates, which were then used to simulate plasma concentration-time profiles. Predicted plasma concentrations were subjected to 10 % residual error to include real-world variability. A total of 1000 patients were tested.

The resultant valproic acid plasma concentration-time profiles (Figure 1.6) were analyzed to quantify the effects of the missed doses and hepatic enzyme induction in terms of the change in steady-state maximum and minimum concentrations (C_{max} and C_{min}, respectively). The simulated profiles illustrate the dramatic fluctuations in valproic acid plasma levels that accompany one or two missed doses and the three temporary dosing regimen adjustments. Although there was no actual link to the PD outcomes, the authors assumed that the lower limit for clinical efficacy was 50 mg/L, below which the likelihood of a breakthrough seizure is greatly enhanced, and that plasma concentrations greater than 100 mg/L increased the probability of toxicity. Using these criteria, the authors concluded that the probability of inefficacy (i.e., seizing) is greatly enhanced

when as little as one dose is omitted, especially for those taking antiepileptic medications. Uninduced patients are at low to modest risk for toxic events, even when the dose is tripled. Induced patients, however, are likely to experience transient toxicity due to elevated plasma concentrations. Given their observations, Dutta and Reed recommended a modified dosing regimen for non-compliant patients who miss one or two divalproex doses. For patients treated with concomitant antiepileptic medications, an omitted dose should be offset if the individual recalls the miss up to 12 hours later, noting the risk of adverse events. If the same patient recalls the omission 18 or 24 hours later, the patient/clinician is advised against full-dose compensation (i.e., either doubling or tripling the dose depending on the time) due to risk of clinical toxicity. In these instances, the clinician is encouraged to consider a partial dose replacement approach in conjunction with a return to the scheduled dosing regimen. As for uninduced patients, they should replace the dose upon recalling the miss, even if two consecutive doses are omitted. The authors also pointed out that the simulation could be revised to model young pediatric and/or geriatric patients with better estimates of their PK parameters (i.e., population-specific variability).[71]

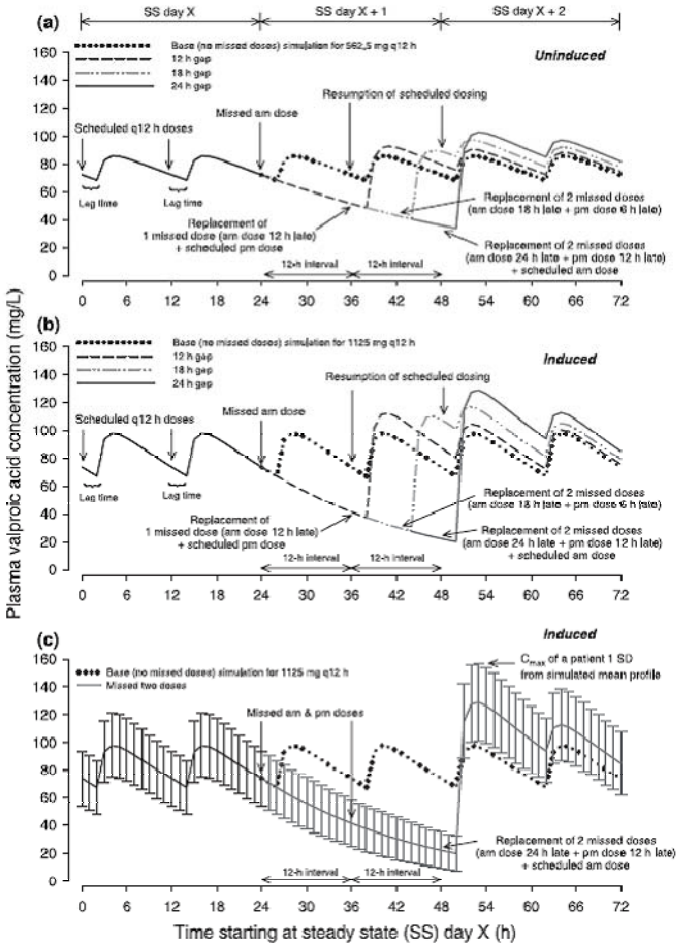

Figure 1.6. Simulated valproic acid (VPA) plasma concentration-time profiles of epileptic patients at steady-state (SS) following the miss of one or two scheduled divalproex doses, with replacement 12, 18, or 24 hours later followed by continuation of scheduled dosing regimen. (a) VPA concentration-time profiles in the mean simulated patient administered 562.5 mg q 12 hours. (b) and (c) VPA concentration-time profiles in the mean simulated patient administered 1125 mg q 12 hours with hepatic enzyme-inducing medication. Panel (c) is the same data shown in (b) with the addition of the inter-patient and residual variabilities in VPA concentration-time curves. The error bars of (c) represent standard deviation (SD). Solid line (0-24 h) to dotted line represents baseline steady-state profile (no missed doses) for the mean simulated patient, and the solid line (24-72 h) represents the predicted mean change after omitting two doses, with replacement 24 h later and resumption of scheduled dosing (triple dose).

Watanalumlerd *et al*. employed MCS and PK modeling to study the effects of gastrointestinal transit on plasma concentration-time profiles of an orally administered combined immediate-release and enteric-coated amphetamine pellet formulation in the fed and fasted states.[72] Compartmental PK were assumed with first-order absorption of the immediate-release pellets, zero-order gastric emptying rate in the fed condition or

first-order gastric emptying rate in the fasted condition of enteric-coated pellets into the intestine, first-order absorption of the drug after being released from the enteric-coated pellets, and first-order elimination. Once absorption occurs, it is assumed that the PK can be described by a one-compartment model. The PK equations included terms for the gastric emptying time as well as the lag time of gastric emptying. The PK parameters for amphetamine (i.e., absorption rate constant, elimination rate constant, apparent volume of distribution, and the fraction of the drug absorbed) were obtained by fitting plasma concentration data obtained from the available literature. Constant doses of 20 and 30 mg were administered for the fed and fasted states, respectively.

Given that the primary focus of this work was to investigate the effects of gastric emptying, only the terms for gastric emptying time and the lag time of gastric emptying were varied during the simulations. The means and standard deviations for gastric emptying time and lag time of emptying were also taken from literature. Separate estimates were available for the fed and fasted states, and a total of four lognormal distributions were generated for the time parameters. Variability of 30 % was implemented for lag time in the fasted state since an estimate was not accessible; this resulted in a range comparable to the other gastrointestinal transit parameters. Each simulated plasma concentration-time curve portrayed the mean of 500 trials, and the variability was captured in one of two ways: (1) by displaying standard deviation error bars or (2) by plotting the 25^{th}, 50^{th}, and 75^{th} percentiles, for each simulated time point. Two lines representing the minimum and maximum simulated concentrations for each time point were also provided.[72]

The accuracy of the simulated plasma concentration-time profiles was assessed by comparing the percent difference between the actual and predicted time to the maximum concentration (C_{max}). A difference of 2.8 % was observed in the fed state, while an 8.4 % difference was observed for the fasted condition. The simulations confirmed the authors' suspicion that the combined formulation, in both the fed and fasted states, would not yield a double-pulsed release profile typical for two immediate-release doses administered at different times because the dosage system remained intact as it passed through the gastrointestinal tract. Rather, the profiles were characteristic of a typical per-oral sustained release formulation. This work confirmed that PK modeling of combined immediate-release and enteric-coated pellets should consider the effects of gastric emptying and gastrointestinal transit.[72]

While the majority of the above examples incorporate several components of the drug development process as they relate to clinical trials, the work of Watanalumlerd *et al*.[72] exemplifies that any given element can be studied independent of its counterparts. Integration is, however, a logical extension of studying each component independently. Simulation is expected to enhance the efficiency of the drug development process by generating a greater understanding of the drug itself and its safety and efficacy within patients. In doing so, the significant cost associated with producing innovative medical products is likely to be reduced. Simulation can also be harnessed to analyze and reduce the intrinsic risk of pharmaceutical products.

1.3.3.3 Probabilistic Risk Assessment

Probabilistic risk assessment is a systematic method to quantitatively characterize the risk of a given system. Assuming the risk is detectable, it consists of two

components, probability and severity. Probability describes the likelihood of occurrence for an adverse event, whereas severity expresses the magnitude of the outcome. When conducting a PRA, risk is expressed through a risk model, a function comprised of the various terms known or thought to impact adverse events. At least one of the inputs to the risk function is described by its probability distribution as opposed to a scalar (e.g., central tendency); methods that only use single values as inputs are referred to as point estimate techniques.[65]

Depending on the underlying objectives, PRA is used to evaluate the variability and/or uncertainty of risk estimates. Variability is defined as the true heterogeneity of a variable or a response for a sample or population, while uncertainty is the error associated with the parameters or models. For example, parameter uncertainty may arise due to questionable model inputs, which may be a consequence of the representativeness of the data for a given sample population. Model uncertainty could be a result of ambiguity in the estimated model coefficients or perhaps the structure of the model used for the risk function (e.g., linear versus nonlinear). It is important to note that uncertainty can be reduced (e.g., acquire additional or improved data), whereas variability is inherent to a population. Variability, however, can be better characterized with more data, but cannot be reduced or eliminated. The output of a given PRA trial is the observed range of probability distributions, which, depending on the input distributions, characterize variability or uncertainty. The results of PRA can be used to better allocate resources (e.g., personnel, finances) and establish performance objectives to mitigate risk. Thus, PRA is an effective means of weighting initial investments based on potential returns.[65]

One of the main objectives when conducting a PRA is to determine the distribution(s) of possible risk scores. This illustrates the range of threats the assessor can expect. It is also essential to understand which factors have the largest impact on risk. This is accomplished through a sensitivity analysis, which is a systematic method to delineate the dependency of the risk estimates on variability or uncertainty in the risk factors. It is used to quantify the relative contribution of each model input to risk and to understand important sources of variability or uncertainty. This is imperative since subtle changes to one factor may have a significant effect on the risk estimate, whereas other factors may be relatively insensitive to large fluctuations.[65]

Sensitivity analyses can range from extremely simple techniques to those that utilize relatively complex mathematical and statistical approaches. The latter are more common since numerous sources of variability and uncertainty tend to simultaneously affect the risk estimate. Simple techniques might include determining the range of possible values or quantifying the percent of total risk for each factor. More complex approaches might include multiple linear regression or some other statistical analysis to assess the percent variance in the risk estimate explained by each factor.[65]

A common metric used in sensitivity analyses is the sensitivity ratio, which is sometimes referred to as the elasticity equation. The sensitivity ratio (SR) is expressed as

$$SR = \frac{\left(\frac{Y_2 - Y_1}{Y_1}\right) \cdot 100\%}{\left(\frac{X_2 - X_1}{X_1}\right) \cdot 100\%} \qquad (1.2)$$

where Y_1 is the baseline output obtained using baseline values for the input variables, X_1 is the baseline point estimate for the input variable X, X_2 is the new value of the input

variable X after the change, and Y_2 is the new output after changing the value of one of the input variables X to X_2. For PRA, the input is a specific value for a given risk factor and the output is the risk estimate. Risk estimates are most sensitive to factors that yield the highest absolute value of the SR and least sensitive to those with the lowest absolute value of the SR. Although popular and instructive, the SR has limited applicability. The SR approach assumes independence of the input variables. Therefore, if two or more inputs interact, varying one input while another is held constant can misrepresent the actual impact on risk. When this is the case, confounding variables must be allowed to vary simultaneously.[65]

Another metric commonly used in sensitivity analyses is the coefficient of determination (R^2). With this method (i.e., simple linear regression), the square of the correlation coefficient (r) between the various input values for a given factor and the result risk estimates is reported; the correlation coefficient itself is powerful because it indicates whether the factor is positively or inversely correlated to risk. Factors with an R^2 close to 1.0 have are highly sensitive whereas those close 0.0 are nearly insensitive. The coefficient of determination for the risk factors can also be represented as the percentage contribution to total variance of risk. Numerous other statistical approaches, such as the Spearman rank correlation and multiple linear regression, are also valid for performing sensitivity analyses. The sensitivity analysis methods ultimately employed will depend on the level (i.e., discrete or continuous) of the input and output variables and the form of the underlying risk model (i.e., linear or nonlinear).[65]

Sensitivity analyses performed on inputs to certain risk models, in particular those that are nonlinear, have the potential to be highly dependent upon the values used to

perturb the model. For that reason, it is beneficial to examine a broad range of values for the inputs to delineate their impact on risk. Inputs characterized as random variables can be described by a specified probability distribution. As such, the inputs can assume values within this distribution and, therefore, can be easily studied using MCS.[65] Likewise, MCS can be used to evaluate the uncertainty in input variables or model parameters.

1.3.3.4 Monte Carlo Simulation and Probabilistic Risk Assessment

Probabilistic risk assessment has long been used as a standalone gauge of adverse events. Wreathall and Nemeth suggest that this technique has (conceivably) been used most frequently within the commercial nuclear power industry;[73] this is not to say, however, that PRA was not used earlier by other industries. Beginning in the early 1970s, the US Atomic Energy Commission sought to estimate the number of accidents that could potentially result in the discharge of radioactive materials. While PRA itself can be used to analyze a system (as can MCS), several guidance documents underscore the utility of employing it with MCS. The National Aeronautics and Space Administration (NASA) details their use of PRA in conjunction with MCS to help direct decisions to ultimately augment safety and program performance.[74] Similarly, the Environmental Protection Agency (EPA) has applied PRA and MCS to analyze ecological risk for the support of risk-based decisions since 1997[75] and proposed equivalent approaches for analyzing risk to human health in 2001.[65] The 2001 EPA Superfund guidance illustrates how to use MCS to estimate exposure and risk, and discusses the role of PRA in their decision making process. The time, resources, and expertise required to effectively perform PRA are noted as drawbacks to this technique.

While this may be true, the potential benefits and savings realized by such a program well outweigh these requirements.

Inputs to PRA risk expressions can assume different values based on their locale within the population; as such, these inputs are said to be random variables. A continuous variable can be described by its associated probability density function or cumulative distribution function, whereas a discrete variable can be characterized by its corresponding probability mass function. These functions describe the probability of the parameter assuming a given value. When defining or selecting the appropriate values or distributions for input parameters, there often is an underlying level of ambiguity regarding these data. For example, in situations where data either do not exist or have yet to be collected, estimates are often drawn from previous assessments or suitable literature. More appropriate data can be incorporated into the PRA if and when they become available. Thus, it is the responsibility of the individual(s) executing the assessment to defend the assumptions that were made and properly communicate their implications and constraints.[65]

Once the risk equation(s) have been defined and the corresponding input distributions have been characterized, MCS can be used to repeatedly extract input parameters at random to evaluate their influence on risk. In other words, each iteration tests a potential real-world scenario (e.g., 80 kg, 65 year old asthmatic male with congestive heart failure dosed with a theophylline tablet coming from a batch that was found to have unacceptable content uniformity) to better understand the risk associated with the particular set of conditions. With the estimated risk distribution, it is then possible to determine if appropriate action is necessary. Results of PRA trials and the

actions that follow are dependent on not only the individual or organization conducting the assessment, but also the rationale behind its initiation. For example, the output may indicate that the underlying distribution for a given parameter (e.g., content uniformity of an active pharmaceutical ingredient) needs to be narrowed to lower the probability of observing toxicity in patients.[64,65]

Risk assessment is frequently used to estimate the potential hazards associated with various substances, including pharmaceutical compounds. For example, molecules that are identified as potentially carcinogenic, mutagenic, and/or teratogenic are typically entered into a risk assessment program to characterize the relative risk between dose and response (e.g., development of a tumor). In order to accomplish this, it is necessary to estimate the probability of the expected outcome as a function of the dose administered. The elucidation of this relationship, however, fails to deconvolve the underlying pharmacokinetic and pharmacodynamic mechanisms that contribute to the undesirable response. Therefore, integration of PK knowledge has been proposed as a means to enhance the power of the risk assessment.[76]

Kodell *et al*. conducted a probabilistic dose-response assessment using tumor dose-response data from a 2-year rodent bioassay to investigate several methods of integrating PK and PD knowledge within a dose-response model.[76] The authors used MCS to study the impact of linking various combinations of PK and PD models on the assessment of risk. Further, MCS was exploited to randomly generate rodent tumor responses based on doses sampled from an assumed PK dose distribution. They concluded that the use of PK data, which related administered dose to *in vivo* levels, reduced the uncertainty associated with assessment of tumor risk. They also determined

the PK models of various complexities did not affect the variability in the risk estimate. This was attributed to the binary model for tumor response (i.e., tumor or no tumor).

Although the study by Buur *et al*. does not involve the administration of a drug with the intent to treat humans, many of the concepts and methods utilized within are directly applicable to the pharmaceutical industry.[77] The authors performed a risk assessment when they used probabilistic modeling and MCS in conjunction with a physiologically-based PK model to predict the withdrawal times of sulfamethazine residues in swine tissue. The FDA regulates the period of time that must elapse before feed animals can be harvested for consumption following the administration of certain agents to mitigate the risk of adverse reactions in humans (i.e., withdrawal time). Sulfamethazine is an antibiotic used for the treatment and prevention of several diseases commonly contracted by pigs. Sulfonamide drugs are of significant interest to regulatory officials since they are known to cause a variety of allergic reactions in humans.

The authors utilized a published physiologically-based PK model and numerous published values for the parameters of interest (e.g., hepatic clearance rates, rate of absorption, percent plasma protein binding). A sensitivity analysis was conducted to determine the parameters that significantly affected the pharmacokinetics of sulfamethazine; specific details regarding the statistical approaches employed were not provided. Lognormal distributions for the statistically significant parameters (inputs to the PK model used to predict tissue and plasma concentration) were generated using the widest dispersion estimates published so as to output the most conservative withdrawal time estimates. These distributions, therefore, represented the variability within the general swine population. Insensitive parameters were not utilized during the subsequent

MCS analyses. Each MCS trial iterated 1000 times to yield a single withdrawal time. The withdrawal time was estimated as the upper limit of the 95th percent confidence interval of the time necessary for 99 % of all tissue and plasma concentrations of the 1000 trials to decrease below the specified threshold. A series of 100 trials were conducted for specific tissues of interest. The validity of the simulation architecture was assessed by comparing the predicted plasma and tissue concentration profiles to published mean concentration-time data sets; it was determined that the predicted concentrations corresponded well with the reported values. The authors concluded that the mandated withdrawal time of 15 days should be revisited in lieu of the considerable public health risk that persists well after this time window. Although the authors acknowledge that their methodology most likely yields more conservative estimates than those that might actually be observed, they underscore the fact that their simulation architecture can be updated as additional data which better approximate the true parameter distributions are acquired.[77]

1.3.4 *In Vitro-In Vivo* Correlation (IVIVC)

1.3.4.1 Background

Pharmacokinetic and pharmacodynamic models are generated using clinical data acquired with the assistance of healthy volunteers and ailing patients. While testing in humans is a critical component of the drug development process for a new drug application (NDA), there is a distinct need for a method capable of assessing the clinical effectiveness and reproducibility of medical products within an artificial environment. This is especially relevant in the case of abbreviated new drug applications (ANDAs) for

generic drugs, where clinical studies to establish safety and effectiveness are generally not required. Knowledge obtained during pharmacokinetic, and to a certain extent pharmacodynamic, modeling is used to help establish a relationship between *in vitro* and *in vivo* performance. Predictive relationships between *in vitro* and *in vivo* product performance decrease development costs, reduce unnecessary burden to test subjects, and expedite product release, all of which underpin public health objectives of the FDA.

In vitro-in vivo correlation is the mathematical architecture for relating *in vitro* drug release profiles to absorption *in vivo*. The main purpose of IVIVC is to demonstrate *in vivo* bioavailability through *in vitro* analyses. Dissolution testing is the conventional *in vitro* test employed by the pharmaceutical industry to assess drug release profiles in view of the fact that drug dissolution and release from the dosage form are acknowledged as key elements to clinical performance; dissolution was formally recognized as a sensitive and reproducible surrogate for assessing bioequivalence in 1993. Dissolution testing supports manufacturing quality control programs, the determination of product release characteristics, and certain regulatory considerations.[22,78,79]

Dissolution testing monitors the extent or rate of drug release as a function of time. The United States Pharmacopeia (USP) defines 4 different apparatus (1 – 4), the basket apparatus, the paddle apparatus, the reciprocating cylinder, and the flow-through cell, for general dissolution testing.[80] For certain compounds, individual monographs dictate the specific requirements for dissolution analyses. The FDA encourages the use of apparatus 1 or 2 for the establishment of an IVIVC; however, apparatus 3 and 4 may be employed when the dissolution properties cannot be ascertained with the former setups. Equally important to the setup is the analytical method utilized to monitor drug

release. Ultraviolet-visible (UV/Vis) spectroscopy or high pressure liquid chromatography (HPLC) are the two most common analytical techniques used to measure API concentration within the dissolution media. Given the overall significance of dissolution testing, the methodology and its associated specifications are often justified through IVIVC efforts. In addition to its role in instituting scale-up and postapproval changes (SUPAC), establishing biowavers, and setting dissolution specifications,[79] IVIVC has also been associated with enhanced product quality.[78]

Clinical data are traditionally available early on in the drug design and development process. IVIVC efforts commence during the initial stages of development and can continue late into the life cycle of a product. Numerous prototype formulations and dosage forms, with various *in vitro* and *in vivo* characteristics, are typically considered during product development. Although the majority of IVIVCs reported in literature are for per oral dosage forms, research is underway to establish correlations for other delivery vehicles (e.g., transdermal patches, biodegradable parenteral depot systems).[78,79] The IVIVC begins by proposing an *in vitro* target to meet a desired *in vivo* performance profile or specification; the targets are subject to change, however, as the product characteristics are finalized to achieve the intended performance. The desired performance profile may also be modified. As more and more data are generated, the IVIVC is refined to accurately reflect the relationship it seeks to describe.

1.3.4.2 Classification of IVIVCs

Numerous approaches to IVIVC modeling are defensible. Moreover, IVIVCs of disparate complexities are suitable depending on the given application. Both the USP and the FDA have released documents intended to help direct participants in constructing

an IVIVC appropriate to their needs. The FDA published a guidance on IVIVC in 1997 entitled *Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations*.[22] Likewise, a chapter titled *In vitro and In vivo Evaluations of Dosage Forms* can be found within the USP.[81] Within these documents, four different levels of IVIVC are acknowledged: A, B, C, and Multiple Level C (certain articles recognize a fifth level, D;[78] however, since it is a rank ordering and, therefore, not a true correlation, level D will not be considered within this work). The levels are defined based on the amount of data used to establish the relationship, which directly determines the interpretability of the IVIVC function.

Level A correlations are the most powerful IVIVC and are advocated by the FDA.[22] As Level A correlations use all *in vitro* and *in vivo* data, they represent a point-to-point relationship between the *in vitro* dissolution response and the *in vivo* input rate, where the latter is typically expressed as the *in vivo* release of the drug from the delivery vehicle. Linear correlations, regardless of whether the *in vitro* and *in vivo* curves are directly superimposable or are rendered so by the implementation of a scaling factor, are the most common Level A correlations. Nonlinear solutions, however, are not incorrect and therefore, should not be overlooked in situations where linear correlations are infeasible. Once the association between the *in vitro* dissolution rate and the *in vivo* input rate is understood, the relationship must be extended to include the portion of drug absorbed *in vivo* (the relationship is not, however, extended to the therapeutic outcome). Ultimately, Level A correlations should be completely predictive of a drug's *in vivo* performance (e.g., plasma drug concentration). This type of correlation can be used to

justify a change in the manufacturing route, site of production, incoming raw materials, and/or for certain minor formulation amendments.

Level B correlations are similar to Level A in that they use all *in vitro* and *in vivo* data; however, they are not defined point-by-point. Rather, Level B correlations are based on statistical moment analyses, particularly the first moment. For Level B correlations, the mean *in vitro* dissolution time is related to the mean *in vivo* residence time or the mean *in vivo* dissolution time. It is critical to note that dissimilar *in vivo* curves can produce comparable mean residence times. Since Level B correlations do not model absolute plasma time curves, they alone cannot be used to justify the same changes that can be addressed with Level A correlations.

Level C correlations relate one dissolution metric (e.g., $t_{50\%}$, $t_{90\%}$) to one pharmacokinetic parameter (e.g., C_{max}, T_{max}, AUC). Considering that a Level C is a single point correlation, it does not communicate the complete *in vivo* plasma profile. Since only a partial relationship between dissolution and absorption is ascertained, a Level C is the weakest of all the correlation levels. The applicability of a Level C correlation is analogous to a Level B correlation. While neither establish bioequivalence, both may be useful in product design, particularly in optimizing formulations.[78]

Multiple Level C correlations relate several dissolution time points, preferably a minimum of three representing initial, intermediate, and ending time values, to one or more pharmacokinetic parameters. Each time point should be related to the same variable when more than one pharmacokinetic parameter is implemented. A multiple Level C correlation established using time points representative of the entire dissolution

profile can serve as evidence for a biowaver. More often than not, a Level A correlation is feasible if a Level C correlation has been established.

1.3.4.3 Approaches to IVIVC Modeling

Deconvolution and convolution are two common methodologies employed to construct IVIVCs. The former is a two-stage modeling technique, whereas the latter involves a single stage. The first stage of deconvolution entails modeling of the *in vivo* absorption or dissolution time profile. This can be accomplished using a number of pharmacokinetic techniques such as the Wagner-Nelson or Loo-Riegelman methods, or general non-compartmental schemes.[82] Once the *in vivo* time profile has been estimated, the second stage of the deconvolution method is to determine the relationship between the *in vivo* profile and the *in vitro* dissolution profile. The goal is to establish a point-to-point relationship between the corresponding *in vitro* and *in vivo* parameters acquired at the same time. This can be done by way of a simple linear relationship or a more intricate sigmoidal (e.g., Hill) function. Deconvolution methods suffer from restricted modeling flexibility (due to the numerous constraints imposed by the methods of stage one) and do not convey drug plasma concentrations (as they model fraction dissolved versus fraction absorbed), which severely limit the interpretability of the IVIVC.[22,83]

Conversely, convolution directly relates the *in vitro* dissolution profile to the *in vivo* drug plasma concentration time profile through a convolution integral

$$C(t) = r(t) * C_{\delta}(t) = \int_0^t C_{\delta}(t-\tau) r(\tau) d\tau \qquad (1.3)$$

where C represents the drug plasma concentration, r is the *in vivo* input rate, τ is the dosing time, t is the current time, and C_{δ} is the instantaneous absorption of a unit quantity

of drug, which can be estimated from IV bolus, oral solution, suspension, or rapidly releasing (*in vivo*) immediate release dosage forms. In other words, total plasma concentration C(t) is the summation (i.e., integral) of the remaining fractions $C_\delta(t-\tau)$ of the infinitesimal contributions $r(\tau)d\tau$ occuring at time τ.[22,83] The fundamental objective of convolution-based IVIVC methods is to determine the functional relationship that connects the *in vivo* input rate (r) to the *in vitro* dissolution rate (r_{dis}) such that

$$r = f\left(r_{dis}, t\right) \tag{1.4}$$

The most parsimonious convolution-based IVIVC model involves a linear relationship

$$r\left(t\right) = a_0 + a_1 \cdot r_{dis}\left(t\right) \tag{1.5}$$

where the scale (a_1) and offset (a_0) coefficients are 1 and 0, respectively. This has been termed the basic convolution-based IVIVC method.[83] While this is an ideal scenario, the scale and offset coefficients may often be different than 1 and 0, respectively. Due to factors that frequently prohibit instantaneous uptake *in vivo* (e.g., time necessary to transport to absorption site) and factors that yield differences in units of measurement (e.g., fraction of the dose absorbed), time and amplitude scaling in addition to lag-time coefficients are often beneficial to model performance, which can be achieved using the equation

$$r\left(t\right) = s_r \cdot r_{dis}\left(t_0 + s_1 \cdot t\right) \tag{1.6}$$

where s_1 and s_r are the time and amplitude scaling factors, respectively, and t_0 is the absorption lag-time term, or the time at which the drug is first absorbed systemically. Scaling and offset coefficients should only be applied in situations where they can be justified mechanistically and enhance the predictive power of the model.[83]

While deconvolution- and convolution-based methods have been applied for some time,[22] a novel method known as direct, differential-equation-based IVIVC was proposed by Peter Buchwald in 2003.[83] Although the direct, differential-equation-based IVIVC method is analogous in many ways to the convolution technique, it is unique in that it harnesses differential equations, which are linked to a basic pharmacokinetic model that describes the system under consideration. Plasma concentration, therefore, is mathematically expressed by a functional pharmacokinetic model that describes the effect of the body on the drug itself. The assumption is made that absorption, distribution, and elimination of the drug can best be modeled using compartmentalized pharmacokinetics. For ease of demonstration, the following discussion will assume a one-compartment model; however, multi-compartment models can be used during differential-equation-based IVIVC.

The fundamental architecture of differential-equation-based IVIVC expresses *in vivo* drug plasma concentration as

$$\frac{dC}{dt} = r(t) - kC \tag{1.7}$$

where k represents the elimination rate constant, and C, r, and t are as were previously defined. In terms of pharmacokinetic modeling, the *in vivo* input rate (concentration per time) for a one-compartment open model with, for example, exponentially decreasing input with an absorption rate κ, can be expressed using the common formulae

$$r(t) = \frac{\kappa D_0}{V} \cdot e^{-\kappa t} \Rightarrow C = \frac{\kappa D_0 \kappa}{\kappa V \kappa - k} \cdot \left(e^{-kt} - e^{\kappa t}\right) \tag{1.8}$$

where D_0 represents the initial dose and V is the volume term. For the purposes of IVIVC, however, the *in vivo* input rate (r) is related to r_{dis} through a function analogous to

that used in the convolution method (Equation 1.4). This assumes, however, that dissolution, not absorption, is the rate limiting step.

Buchwald[83] proposed an amended model that includes φ_{abs}, a time-dependent absorption factor that takes into consideration fluctuations in absorbance as the dosage system traverses the gastrointestinal tract

$$r(t) = \varphi_{abs}(t) \cdot s_r \cdot r_{dis}(t_0 + s_1 \cdot t) \tag{1.9}$$

The time-dependent absorption factor can be as straightforward as a low-pass filter ($\varphi_{abs} = 1$ if $t \leq t_{cut}$; $\varphi_{abs} = 0$ if $t > t_{cut}$) or can be something more complex such as a sigmoidal step-down function

$$\varphi_{abs} = \frac{e^{-\eta(t-t_{cut})}}{1+e^{-\eta(t-t_{cut})}} \tag{1.10}$$

where η is the steepness of the cut-off and t_{cut} is the last time point at which absorption occurs. To minimize the number of parameters that must be optimized, η can be set to a constant value. Due to its extensive surface area, the small intestine is the primary site of absorption for many drugs. The mean small intestine transit time has been reported to be 3 (±1) hours.[84-87] Taking into consideration the range of gastric emptying times, minimal systemic uptake would be expected beyond 4 to 10 hours post administration; drugs that absorb well within the colon (i.e., certain highly-permeable molecules) would extend the aforementioned absorption window. Thus, the application of a time-dependent absorption factor is justifiable for many controlled release delivery systems. Buchwald reported enhanced IVIVC performance when a sigmoidal absorption function ($t_{cut} = 6.4$ hours) was used in conjunction with scaling and offset factors.[83]

A total of two equations are necessary to establish an IVIVC for a one-compartment pharmacokinetic model (Equations 1.7 and 1.9). Each additional

compartment incorporated in the pharmacokinetic model will require one additional differential equation to describe the change in concentration as a function of time within that compartment. Estimations of r_{dis} at numerous time points are required to numerically solve the IVIVC equations. The dissolution rate can be estimated by fitting the *in vitro* dissolution profiles with a flexible, continuous function, such as a sigmoidal, quadratic, or exponential (e.g., Weibull) expression. The instantaneous dissolution rates are then acquired via the analytical derivative of the fitted function.[83]

The direct, differential-equation-based IVIVC method is restrictive in that it assumes compartmental-based pharmacokinetics. It does, however, yield an estimation of the elimination rate constant (k), which can be verified for accuracy using clinical data. Moreover, it circumvents the need to convolve or deconvolve mathematical expressions, and it directly relates the *in vitro* dissolution profile to the *in vivo* drug plasma concentration time profile.

1.3.4.4 Predictability of IVIVCs

The validity of an IVIVC is demonstrated by its predictability. Recalling that the key objective of an IVIVC is to generate a predictive mathematical function that relates *in vitro* and *in vivo* performance, validity is centered on the degree of prediction error. Validity can be assessed via the data used to construct the IVIVC (referred to as internal predictability) or data independent of those used to generate the model (referred to as external predictability). Predictability assessments via internal and/or external methods are dependent on the intended use of the IVIVC and the therapeutic index of the drug; estimations of internal variability should be performed for all IVIVCs. Narrow therapeutic index drugs require external estimations of prediction error. Greater detail

regarding the requirements of internal versus external predictability can be found in the FDA guidance.[22] Regardless of the method used, predictability must demonstrate that the IVIVC model can accurately predict *in vivo* performance from *in vitro* data comprised of various release rates and manufacturing conditions. Generally, this should be accomplished using a minimum of three different formulations with release rates that vary by, for example, 10 % for each formulation, which should yield a commensurate distinction in *in vivo* performance. Prediction error can be assessed as the average absolute prediction error (% PE) and typically must be less than 10 % to demonstrate predictability for both internal and external estimations. An adequate number of subjects should be studied to effectively characterize *in vivo* performance. This has been accomplished with as few as 6 and as many as 36 individuals, but this is not to say that more cannot be evaluated.[22] Jaber Emami conducted a thorough review of IVIVC in 2006 and noted that the majority of the literature articles failed to report (or perhaps even assess) predictability.[78]

Dissolution is often used as a quality control gauge to assess batch-to-batch or lot-to-lot similarity/dissimilarity. The variability of release at a given time point(s) is commonly assessed, where the acceptance specification could be set as ± 10 % deviation from the average profile of the clinical/bioavailability samples. The power of the analysis is amplified if a predictive IVIVC has been established. Given that dissolution now conveys *in vivo* performance, the dissolution release specifications can be established with aid from the IVIVC to mitigate *in vivo* variability.[22,79]

1.3.4.5 IVIVC for Theophylline Dosage Systems

Theophylline is categorized as a class I compound according to the Biopharmaceutics Classification System (BCS).[88] As a class I molecule, theophylline exhibits high solubility and high permeability. Drug dissolution, not absorption, is the rate-limiting step to the bioavailability of theophylline;[89] when dissolution occurs rapidly, however, the rate of absorption is then determined by the frequency of gastric emptying. Therefore, an IVIVC is not unexpected for theophylline dosage systems, especially for controlled release products where dissolution, not absorption, is the rate-limiting step.

El-Yazigi and Sawchuk remarked that even though the bioavailability of theophylline had been thoroughly investigated, the potential of establishing a quantitative IVIVC had yet to be reported as of 1985.[90] This finding was the impetus for investigating the effects of pH, apparatus, and stirring speed on the dissolution rate of theophylline from assorted commercially available products. In doing so, the authors modeled the cumulative percent theophylline dissolved using a first-order equation. The covariation between the cumulative percent dissolved at various times or the first-order rate constant was assessed against different pharmacokinetic parameters (e.g., the fraction of the dose absorbed at a given time, dose-normalized area under the curve, peak serum levels). Pharmacokinetic parameters were estimated from the *in vivo* data, which were collected in male New Zealand white rabbits. At the time of publication, a coefficient of determination of 0.9 or greater backed by reproducible predictions of the modeled bioavailability parameters signified a good IVIVC. The authors reported strong Level C correlations amongst *in vitro* and *in vivo* parameters for the dosage forms assessed under a myriad of experimental conditions. Moreover, good agreement between the actual and

predicted pharmacokinetic parameters was reported (for one apparatus at a specified pH and a constant stirring rate).[90]

Munday and Fassihi generated IVIVCs for the commercially available controlled release product Theo-Dur as well as their own novel controlled release delivery system Mintab.[89] All *in vivo* data was collected in canines. The authors used the rudimentary deconvolution-based approach to IVIVC modeling where a one-to-one correspondence between the *in vitro* and *in vivo* data is desired. Although the Theo-Dur tablets demonstrated good IVIVC, poor correlation between the *in vitro* dissolution and *in vivo* absorption of the Mintab system was observed. This was attributed to the incongruent dissolution profiles observed across the various dissolution conditions. The authors did not investigate alternative approaches to modeling the Mintab data. Moreover, they neglected to comment on the predictability of the IVIVCs.[89]

Yu *et al*. investigated the *in vitro* and *in vivo* characteristics of four experimental oral controlled release theophylline dosage systems (three different hard gelatin capsule formulations and one tablet).[91] Healthy male beagles were used to collect the *in vivo* data. The *in vitro* dissolution profiles of the four different delivery vehicles where shown to be unique across the various formulations and dissolution conditions. Using a deconvolution-based IVIVC approach, the authors established a point-to-point (ratio) relationship between the cumulative percent theophylline released *in vitro* and *in vivo*. Although no predictability estimates were determined, the authors concluded that the correlation between the *in vitro* and *in vivo* data was good based on the visual congruence of the point-to-point (ratio) time profiles.[91]

While the previous investigations utilized animals for the *in vivo* analyses, Hussein and Friedman conducted an IVIVC study of various theophylline formulations in both canines and humans.[92] Five experimental sustained release theophylline formulations and two commercial sustained release products (i.e., Theotrim and Theo-Dur) were selected. Two different dissolution media were used to model conditions within the gastrointestinal environment. Distinct mean dissolution profiles were obtained for each dosage system using the specified dissolution methodology. Three male and two female dogs and six healthy humans (gender not indicated) were studied to obtain the *in vivo* absorption data. The percent theophylline absorbed in both humans and canines was estimated using the Wagner-Nelson method. Mean *in vitro* release and individual *in vivo* absorption profiles were used in conjunction with linear regression to relate the percentage released to the percentage absorbed at each time point sampled. The slopes (± standard deviation) and the coefficients of determination for three of the experimental formulations and both commercial products were reported; the remaining two experimental formulations were only tested *in vitro* (no explanation provided). Based on the variance explained through linear regression, the authors concluded that IVIVC was feasible using bioavailability data obtained both in canines and humans. No estimates of prediction error were reported.[92]

1.3.5 Pharmacokinetics

1.3.5.1 Background

The study and understanding of *in vivo* drug (and metabolite) levels over the duration of treatment are key functions in the drug development process. These

concentrations, in turn, influence the clinical outcomes of the therapy and, therefore, must be managed so as to be both safe and effective. Aside from the intricacies of drug delivery systems, numerous variables, including behavioral, environmental, physiologic, and pathophysiologic factors, can influence the uptake and disposition of medical substances. For example, concomitant drug therapies via drug-drug interactions commonly modify the rate of removal of one or several of the drugs from the body, which has the potential to drastically affect patient well-being. Furthermore, many drugs perform differently when administered to infant, adolescent, adult, or geriatric patients due to changes in organ size and function. As such, extensive studies must be undertaken to identify these factors, as well as those associated with the intricacies of the delivery system, and appreciate how (mechanistically) they alter the management of disease. More often than not, these studies yield quantitative empirical models that predict drug serum/plasma levels for groups of patients under a specific set of conditions.

Pharmacokinetics denotes the examination of the time course of drug absorption, distribution, metabolism, and excretion (ADME) to ultimately elucidate the relationship between dose and exposure. Pharmacokinetics can be grossly divided into two components, an experimental and a theoretical element. The former involves the collection and analyses of data, and the methods describing these practices, while the latter entails the generation and validation of physiologically-based or empirical models that express drug levels following administration. The pharmacokinetic focus of this dissertation is predominantly on the theoretical component. Thus, the background and application of experimental pharmacokinetics, which includes topics such as sampling of biological fluids/tissues and the analytical methodologies for measuring drug

concentration, will not be addressed. This work will, however, model plasma rather than serum concentrations given that plasma has the ability to perfuse through all tissues in the body. If a dynamic equilibrium is assumed to exist between plasma and the body tissues, changes in drug plasma concentration thereby mirror those in the tissues.[82,93]

One of the major outcomes of experimental pharmacokinetics that subsequently becomes the impetus for theoretical modeling is the drug concentration time profile, which is an illustration of the concentration of drug in the serum/plasma determined at precise sampling time points following the administration of one or more doses (Figure 1.7). Drug concentration time profiles reflect the substance's absorption into systemic circulation, distribution to the various tissues in the body, and elimination via biotransformation and/or excretion, all of which take place simultaneously during treatment. Countless variables affect the shape of these profiles, including the route of administration, the dosing intervals, the amount of drug administered, the rate of gastric emptying, and the fitness of the individual and the clearing organ(s). The profiles, or more precisely, the data upon which the profiles were generated, are then used as inputs to various numerical models that seek to accurately express the change in drug concentration as a function of time. Otherwise, they can be described by scalar metrics such as the maximum concentration (C_{max}), the time necessary to reach the maximum concentration (T_{max}), the area under the curve (AUC), and/or the drug clearance rate, all of which have particular PK significance.[82,93]

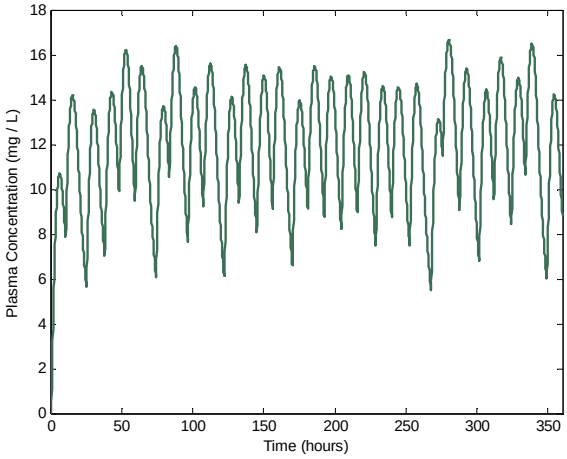

Figure 1.7. Plasma concentration time profile for a dose administered every 12 hours for 15 days simulated *in silico*.

The underlying PK are typically described as liner or nonlinear. Linear kinetics are characteristic for a drug whose PK parameters are invariant with the administration of different doses. For example, first-order kinetic equations might adequately describe the time course of drug ADME; a plot of the AUC for the plasma concentration-time curve versus dose would be linear. Some drugs, however, display dose-dependent (i.e., nonlinear) PK, and the same AUC versus dose plot would be nonlinear over certain ranges. Many of the physiological processes responsible for ADME are conducted by carrier-mediated or enzymatic systems, both of which are subject to saturation. Additionally, the drug (or other concomitant drugs) may induce a physiologic or pathophysiologic change that transforms the kinetics of the drug of interest from linear to nonlinear. These situations are exemplified by a change in the rate of drug elimination. For instance, the elimination rate may decrease due to the saturation of metabolic enzymes or may increase due to the induction of metabolic enzymes. Regardless of the

directionality, dramatic inaccuracies in the plasma levels may occur if the change is inaccurately modeled.[82]

As stated, pharmacokinetic models can be empirical or physiologically-based. Empirical models are applied to a set of data based solely on their ability to mathematically describe the information, regardless of the underlying physiological and pathophysiological mechanics. Alternatively, physiologically-based pharmacokinetic models encompass the physical and biochemical functions of the body with respect to ADME of drugs and metabolites. These models may incorporate knowledge regarding the rate of blood perfusion for a particular organ or the various types of cells that constitute the major organ systems that interact with the drug substance.[82,93]

In theory, an infinite number of models can be ascribed to the complex kinetic processes of ADME if they are scrutinized on a cellular or sub-cellular level. Approaches to modeling, however, normally employ a number of assumptions to simplify these processes while concurrently retaining physiological relevance or numerical applicability. The two most notable categories of pharmacokinetic modeling are compartmental and non-compartmental models. Compartmental models reduce the body down into a number of systematic or serially-related compartments that reversibly interact with one another. Compartments are hypothetical regions that represent a tissue, an organ, or a collection of tissues/organs that display comparable blood flow and affinity for a given drug. Compartmental models assume that the drug is rapidly mixed and is homogenously distributed within a compartment of a definite volume. Moreover, each drug molecule has the same likelihood of exiting the compartment. Rate constants are used to express the transfer of drug between compartments and differential equations

describe the change in drug concentration over time. The instantaneous concentration of drug in the body is, therefore, the summation of concentrations in each of the contributing compartments. Compartment models are open in view of the fact that drug can be eliminated from the system.[82,93]

There are two general forms of compartmental pharmacokinetics. The first is known as the catenary system, which is structured as a series of coupled compartments (analogous to box cars on a train) where input occurs in the first compartment and drug can only be transferred to and from adjacent compartments. Since this format does not effectively characterize the manner in which plasma interacts with tissues/organs, catenary models are used infrequently. The other form of compartmental pharmacokinetics is referred to as the mammillary system, which consists of, at a minimum, one central compartment. They can also incorporate numerous peripheral compartments, all of which have direct (and potentially coincident) access to the central compartment. Mammillary models are the most common structure for compartmental pharmacokinetics.[82,93]

Non-compartmental pharmacokinetics makes no assumption regarding the nature of the distribution of a drug, whereas compartmental modeling presumes that substances are distributed amongst one or more compartments. Non-compartmental PK makes use of statistical moments (e.g., area under the moment curve (AUMC), AUC), which are mathematical descriptions of a discrete distribution of data. A statistical moment of concentration-time data describes the probability density function, which represents the true relationship between concentration and time. A non-compartmental approach to modeling is preferred by some as it is not centered on the same assumptions as

compartmental pharmacokinetics, which traditionally applies empirical models that have no physiologic relevance. Non-compartmental modeling can be of particular benefit in situations where a single compartment model fails to adequately characterize an entire population (e.g., a fraction of the patients are characterized by a two-compartment model, while the remainder are more accurately described by a three-compartment model). Nevertheless, non-compartmental models use differential equations to express the change in concentration (in the body as a whole) over time.

The following discussion and accompanying formulae are a straightforward example of a one-compartment open model. While more complex functions, some of which are nonlinear (e.g., Michaelis-Menten kinetics), exist, the reader is referred to any advanced pharmacokinetic textbook for greater detail. One-compartment open models are particularly useful for drugs administered via IV bolus injection, where the drug is introduced into systemic circulation (nearly) instantaneously. A one-compartment model following IV bolus administration is fundamentally characterized by three terms, the drug in the body (D_B), the apparent volume of distribution (V_D), and the elimination rate constant (k). The apparent volume of distribution is the volume in which the drug is assumed to be homogeneously distributed, and the elimination rate is the collective total of the joint processes of biotransformation and excretion. The apparent volume of distribution does not correspond to a true anatomic space; rather, it signifies the volume of the sampling compartment used to estimate the amount of drug in the body, hence the annex "apparent." Since the human body is more or less at constant volume, it is common practice to use an invariable estimate for the V_D. However, certain physiologic

conditions, such as edema, which results in increases to total body water, can invalidate this assumption.[82,93]

The differential equation that describes a one-compartment model with instantaneous uptake is

$$\frac{dD_B}{dt} = -kD_B \tag{1.11}$$

where D_B and k are as previously defined. Following integration, the drug in the body at any time (t) is expressed as

$$D_B = D_B^0 e^{-kt} \tag{1.12}$$

where D_B^0 is the amount of drug in the body at time t = 0. Mathematically, D_B is estimated as

$$D_B = V_D C_p \tag{1.13}$$

where C_p represents the concentration of drug in the plasma. Given these relationships, the first-order decrease in plasma drug concentration is expressed as

$$C_p = C_p^0 e^{-kt} \tag{1.14}$$

where C_p^0 is the concentration of drug in the plasma at time t = 0 once the drug has equilibrated within the body. The V_D is estimated using the expression

$$V_D = \frac{dose}{C_p^0} = \frac{D_B^0}{C_p^0} \tag{1.15}$$

where the dose is administered via IV bolus injection and C_p^0 is estimated by extrapolating the plasma concentration versus time curve back to the y axis.[82,93]

Clearance is the pharmacokinetic term for the overall rate of removal of drug from the body. It does not identify the underlying elimination mechanism(s); rather,

clearance is purely a numerical expression of the volume of fluid containing drug that is removed from the body per unit time. Although clearance is typically expressed with respect to volume, removal can also be conveyed in terms of mass, fraction, or concentration per unit time. Many drugs are removed from the body via a first-order process (i.e., the rate of removal is only dependent upon the plasma concentration). When expressed as volume per unit time, clearance is assumed to be constant for a first-order process, which is a direct result of assuming a constant V_D; just as was true for V_D, certain physiologic circumstances can invalidate this assumption. As such, clearance (Cl) from the body is expressed as

$$Cl = -kV_D \tag{1.16}$$

where k is the first-order elimination rate constant. The negative sign indicates that drug is being removed from the body. Although the instantaneous rate of drug removal from the body (i.e., amount per unit time) will decrease as concentration declines, clearance (i.e., volume per unit time) will remain constant so long as elimination is characterized by first-order kinetics. The concept of clearance can be incorporated within the expression of plasma concentration for a one-compartment open model to yield

$$C_p = \left(D_0 / V_D\right) e^{-(Cl/V_D)t} \tag{1.17}$$

where D_0 denotes the initial IV dose. Analogous equations exist for alternative routes of administration and for multi-compartment models. Moreover, the concept of clearance is also applicable in non-compartmental pharmacokinetics.[82,93] Extensive discussion regarding the estimation of clearance (e.g., renal excretion and biotransformation rate constants) can be found in pharmacokinetic textbooks and the literature. The reader is referred to these general references for greater detail.

1.3.5.2 Pharmacokinetic Models of Theophylline

Accurate dosing of theophylline has remained a significant challenge since its first applications in the early 20th century. Imprecision can largely be attributed the extensive inter-patient variability for the PK of theophylline. Consequently, substantial effort has been devoted to better understanding the underlying relationship between dose and exposure.

Ginsberg *et al*. investigated the utility of a physiologically-based PK model for discriminating PK differences in neonates and adults. This study was performed using two model drugs, caffeine and theophylline, both of which are know to have disparate PK in neonates and adults. Both drugs were selected since caffeine demonstrates a more dramatic difference as a function of age, despite the fact that caffeine and theophylline are eliminated via similar pathways (refer to section 1.3.2.2). Since the underlying PK mechanisms and the significant confounding factors are, by and large, poorly established for neonates when compared to adults, the authors were particularly interested in investigating the transferability of *in vitro* metabolic parameters determined in mammalian cells transfected with CYP c-DNAs to whole liver metrics. This is an attractive method since the whole liver metrics can then be adjusted for the differential expression levels observed between neonates and adults.

To accomplish this objective, a 5-compartment physiologically-based PK model was generated (i.e., liver, kidney, fat, and rapidly and slowly perfused tissues). Hepatic metabolism was modeled using published *in vitro* Michaelis-Menten constants (i.e., V_{max} and K_m) for the major metabolic pathways of caffeine and theophylline (i.e., CYP1A1, CYP1A2, CYP2E1); all other transfer equations were linear. The Michaelis-Menten

constants were scaled to account for the size of the liver, the amount of microsomal protein per gram of liver tissue, the aggregate concentration of CYP enzymes per microgram of microsomal protein isolated from the tissue, and the relative amount of each CYP enzyme to the total microsomal CYP quantity for a specific age range. Metabolism in organs other than the liver was not modeled and the fate of the metabolites, with the exception of conversion of theophylline to caffeine in neonates (assumed to be first-order), was not considered. Renal elimination was expressed as a first-order process. Although several of the model coefficients (i.e., the Michaelis-Menten constants) were adjusted from the published values to optimize model performance, the authors concluded that the model described the differential PK reasonably well based on the comparison of predicted metabolite and drug levels with urinary excretion data. Adjustments were deemed necessary since the *in vitro* system did not mimic the compensatory pathways that are present *in vivo* (e.g, conversion of theophylline to caffeine, no incorporation of FMO). The authors also noted that metabolism in newborns should be adjusted for gestational variations and postnatal age during the first weeks of life. This study underscores the importance of neonatal PK data. Scaling of theophylline data from adults to newborns would overlook the conversion of theophylline to caffeine that takes place in this age group.[28]

Bjorkman exercised a similar set of objectives when he generated a generalized physiologically-based PK model applicable across a broad range of ages.[30] Unlike the work of Ginsberg *et al*.,[28] however, all relevant PK parameters (i.e., model inputs) were scaled using data obtained from adults as opposed to estimating the parameters from *in vitro* metabolic data. The physiologically-based model was evaluated with theophylline

and midazolam, two drugs that display dissimilar physiochemical and PK attributes (e.g., lipophilicity, predominant CYP metabolic enzymes). As the unbound (i.e., free) fraction of theophylline is known to vary with age, a constant unbound fraction of 0.56 was assumed for adults, while plasma protein binding for infants and children was modeled according to the age-related variability is serum albumin concentration (equation not given in original article). Total theophylline clearance was partitioned to be 85 % hepatic and 15 % renal, and clearance was assumed to be linear (i.e., first-order) across all age ranges studied. For subjects from 0 – 9 years, a bi-exponential growth function was used to assign the relative contributions of CYP enzymes to the hepatic metabolism of theophylline, while 92 % was attributed to CYP1A2 and 8 % to CYP2E1 for individuals 10 years or older. Renal clearance for infants of 6-months was increased by 10 % owing to the methylation of theophylline to caffeine.

The model was validated using amassed literature data. Model performance was assessed according to the percent prediction error for the estimation of clearance (Cl), volume of distribution at steady state (V_{dss}), and terminal half-life ($t_{1/2}$). The median prediction errors for Cl, V_{dss}, and $t_{1/2}$ of theophylline were -4.0 %, 3.4 %, and 24 %, respectively. Bjorkman concluded that the model predicted V_{dss} and Cl well, but noted that inter-subject variability of the actual clearance data was considerable. This is not surprising given the findings of studies such as Jusko *et al*.[46] and Chiou *et al*.[37] The error of prediction for $t_{1/2}$ was appreciably larger; inaccuracies in Cl and V_{dss}, however, are compounded in half-life estimations. Additionally, Bjorkman questioned the legitimacy of several reported half-life values cited for neonates and infants, which may be more inexact than the predicted values. While clearance, and to a much lesser extent, terminal

half-life of theophylline changed as a function of age, the volume of distribution was relatively invariant.[30]

While these examples are far more complex than a linear one-compartment open model, several studies have demonstrated reasonable predictive power using this technique. Many of the following articles have been previously addressed in literature survey (refer to section 1.3.2) and, therefore, will not be expounded upon other than to specify the PK modeling assumptions that were implemented. The studies conducted by Powell *et al*.[45] and Chiou *et al*.[37] utilized a linear one-compartment open model to describe theophylline concentrations. Powell *et al*. also assumed that clearance and volume of distribution were constant for any given individual regardless of the theophylline level. The authors defended their assumptions by referencing the empirical observations of a few studies published prior to their work. The article by Gilman *et al*.,[56] which compared the methods of Powel *et al*.[45] and Jusko *et al*.,[46] and a weighted least-squares Bayesian approach and the method of Chiou *et al*.[37] for the estimation of theophylline clearance, employed a one-compartment open model and assumed linear elimination of theophylline. The authors justified the use of linear elimination based on the several reports that theophylline failed to demonstrate saturable elimination in adults. The mean age of the patients utilized in the Gilman *et al*. study was 43.5 ± 15.8 years. A constant volume of distribution of 0.5 L/kg was implemented.

Brocks *et al*. performed a PK study in 34 pediatric patients ranging in age from 4 months to 14 years of age using a one-compartment open model for orally administered theophylline.[94] The authors generated predicted theophylline serum concentrations using both the patient's individualized volume of distribution (0.3 – 1.54 L/kg) and a

standardized value of 0.5 L/kg, and two blood samples. The mean prediction errors between the observed and predicted serum concentrations using the individualized and standardized volumes of distribution were -0.509 mg/L and 1.27 mg/L, respectively. The authors concluded that their method can be used to accurately estimate theophylline dosages for pediatric patients, despite the fact that it does not require IV data.

Casner *et al*. conducted a randomized trial during which the accuracy of a PK model for IV theophylline dosing was compared to physician dosing.[95] A total of 35 asthmatic or COPD patients were followed throughout the trial (several were excluded due to incomplete data) in one of two randomized groups; the kinetic group was dosed according to a computerized version of the Chiou *et al*. method,[37] while the empirical group was dosed by physicians instructed to obtain a target concentration of 15 mg/L. The physiologic and pathophysiologic characteristics of the two groups were comparable. Three serum theophylline levels were determined from blood samples for each patient. Prediction error for the two groups was estimated by subtracting the third serum level from the target concentration. The mean absolute values were 14.8 ± 4.4 and 12.6 ± 4.1 mg/L for the kinetic and empirical groups, respectively. Despite the fact that the computer predicted dosing was closer to the target value, the difference was not statistically significant. Moreover, none of the clinical outcomes (e.g., number of subtherapeutic or toxic levels, duration of time in hospital) were statistically different between groups. The authors concluded that PK model for theophylline dosing was of no additional clinical benefit. It is noteworthy, however, that a linear, one-compartment PK model was able to match the skill of trained physicians.

The investigation of Hurley *et al*. was similar to the preceding study of Casner *et al*.; Hurley *et al*. conducted a randomized trial to assess the clinical differences of computerized versus physician dosing in 91 asthmatic or COPD patients upon admission to the hospital with acute air-flow obstruction.[53] The authors also implemented a one-compartment open model; however, the physicians were not advised to attain a target theophylline level. Patients were initially dosed according to total body weight and then their doses were adjusted either by a computerized model (i.e., monitored group) or by a physician (i.e., control group). Hurley *et al*. observed no statistically significant difference in theophylline serum concentrations between the two groups, nor in the number of subtherapeutic or superpotent levels. They did, however, observe a statistically significant difference in the lower number of reported subjective side effects (i.e., breathlessness, palpitations) in the kinetic (or monitored) group as well as a statistically shorter hospital stay in the kinetic group (6.3 days) compared to the empirical (or control) groups (8.7 days). Unlike Casner *et al*., the authors concluded that "using a pharmacokinetic method to determine theophylline dosage for the patient with acute air-flow obstruction improves the likelihood of achieving a theophylline concentration in the therapeutic range, and may hasten the patient's recovery."

These are but a few of the examples of the PK models that have been generated to study theophylline. Other researchers have also advocated the use of linear, one-compartment models for the analyses of theophylline delivery systems.[32,35,41,51,52,96,97] Such models are purported to be applicable for both IV and orally administered dosage systems.

1.3.6 Pharmacodynamics

1.3.6.1 Background

A thorough understanding of the physiochemical properties of a drug and how they control its ADME is imperative. Such knowledge, however, offers little insight as to how the molecule interacts with the targeted pharmacophore to elicit a physiologic response. Although a mechanistic understanding of the manner in which a drug docks with the molecular target is not essential early in the development process, a sponsor must quickly identify how patients respond to the drug at a specified dose. Likewise, it is important to understand how the safety and effectiveness of a drug delivery system change as it is optimized with respect to a desired exposure-response profile. Within this, it is important to recognize that optimization will most likely affect the onset, magnitude, and duration of drug action.

Once the pharmacokinetics of a drug are reasonably well understood, it is important to define the minimum effective concentration (MEC) that results in the desired endpoint and the minimum toxic concentration (MTC) that results in any adverse (i.e., toxic) event; concentrations between the MEC and MTC delineate the therapeutic window. Numerous drugs exhibit a proportional relationship between the administered dose and the observed outcome. Many individuals, however, respond differently (although not necessarily unfavorably) to the same drug administered at equivalent doses and it is this inter-individual variability that obfuscates the underlying exposure-response relationship. Moreover, small fractions of the population can respond negatively to a given substance at a specified dose, despite the fact that the majority of individuals tolerate the treatment. It is therefore important to integrate pre-clinical data with clinical

observations, as they become available, to better understand the relationship between exposure and pharmacologic outcome. A thorough appreciation of the exposure-response relationship can help to better pinpoint successful management practices for individual patients or sub-populations *a priori*.[48]

Pharmacodynamics is the field of study devoted to the onset, intensity, and duration of drug action following exposure. In other words, PD is a supplement to PK that draws a parallel between the drug concentration time profile and pharmacologic endpoints to ultimately elucidate the relationship between exposure and response. This bridge between PK and PD helps define the optimum dosing regimen to achieve the intended result. It follows, therefore, that the same intricacies that affect the PK of a drug have the potential to influence the PD. More importantly, these variables can act independently on the PD, despite the fact they showed no observable effect on the PK. Thus, it is necessary to investigate the potential demographic, physiologic, and pathologic factors that affect the exposure-response relationship.[48]

Pharmacodynamic research often involves extensive investigations into the genetic factors that predispose individuals to respond favorably or unfavorably to a given drug. Even in instances where the response is efficacious *per se*, pharmacologic outcomes that deviate from the projected exposure-response profiles are sometimes attributable to genetic polymorphisims of genes that encode receptors specific to the drugs and/or metabolites. Such genetic variations have been known to alter the response independent of any change in the PK curve.[48]

Adding to the significant challenge of defining an exposure-response profile is tolerance. In certain cases, the observed pharmacologic outcome is modulated as a

function of time, despite maintaining the same dosing regimen. This can be attributed to several physiologic or pathophysiologic factors, including an increase or decrease in the quantity of receptors, or an alteration in receptor affinity or signal transduction. Thus, the exposure-response profile can be severely misconstrued if these time-dependent effects are overlooked.[48]

Numerous other phenomena mask the underlying relationship between the PK and PD of a particular drug. For instance, certain molecules elicit no effect below the MEC, while others yield a response distinctive from that observed within the therapeutic window; these effects are thereby exacerbated by sub-potent dosing and/or patient incompliance. In such instances, a single exposure-response curve inadequately characterizes the causal relationship. Furthermore, although many exposure-response curves are linear (or can be transformed to be so) over extended ranges, others are more appropriately described by an alternate function (e.g., sigmoid) due to disproportionate increases or decreases in response at extreme concentrations. Selection of an inappropriate mathematical function to characterize the exposure response relationship can yield an erroneous PD model.

The ability to ascertain the true exposure-response relationship is dependent upon the accuracy of the response estimate. For example, subjective endpoints (e.g., decrease in pain) have higher degrees of uncertainty than measurable responses (e.g., decrease in blood pressure). Responses that are dependent on an individual's or a physician's perception can complicate efforts to generate an accurate PD curve. Further, certain pharmacologic endpoints are not feasible to measure. Therefore, to increase the practicality or lower the associated risk, surrogate PD endpoints are employed under the

assumption that they are predictive of the actual clinical response. For example, analyses of the antitumor effects of chemotherapeutic agents would require routine biopsies, which pose too high of a risk to the patient. To circumvent this risk, clinicians often monitor expression levels of white blood cells to gauge patient response.[98] Use of surrogate endpoints requires an additional level of validation to ensure that they are indeed representative of the true pharmacologic outcome. Regardless of the endpoint employed, the success of any therapy is dependent upon, along with other critical links, a precise relationship between exposure and response.

1.3.6.2 Pharmacodynamic Models of Theophylline

While numerous studies have focused on understanding the complex relationship between theophylline dose and exposure, less research has been devoted to developing pharmacodynamic models expressing the link between exposure and response. This may be, to some extent, justified by the fact that researchers generally agree upon the therapeutic window that characterizes theophylline and the inefficacious and toxic events that occur outside this range. Consistently dosing within this range, however, has proven to be a challenge.

The pharmacologic outcomes of theophylline closely parallel serum or plasma concentrations. Both the degree of bronchodilation and the decrease in airway responsiveness trend with theophylline concentration.[25] In fact, bronchodilation increases linearly with logarithmic increases in theophylline concentration, within the therapeutic range.[35] Over the years of treatment with theophylline, researchers and clinicians have concluded that theophylline concentrations of 10 – 20 mg/L are most likely to safely provide clinical benefit, although it should be noted that levels as low as 5

mg/L are efficacious in certain cases. Others, however, have campaigned for the lower limit of efficacy to be established as 5 mg/L (refer to section 1.3.2). Thus, the MEC is not unanimously acknowledged as 5 or 10 mg/L, or some concentration in between.

Concentrations in excess of 20 mg/L compromise the safety of treatment as the likelihood of toxic side effects increase dramatically above this level. This is demonstrated in Figure 1.8 where the toxic outcomes of 50 adult patients treated with theophylline were documented.[39] Unfortunately, the occurrence of more severe side effects is not always preceded by mild toxic events. While certain mild side effects (e.g., headache nausea) may be offset by the potential clinical benefit, the fact that the onset of more severe side effects cannot be accurately predicted has firmly established the MTC at 20 mg/L, although some have suggested this value should be reduced to 15 mg/L (refer to section 1.3.2). Severe side effects are consistently observed at concentrations well in excess of 20 mg/L.

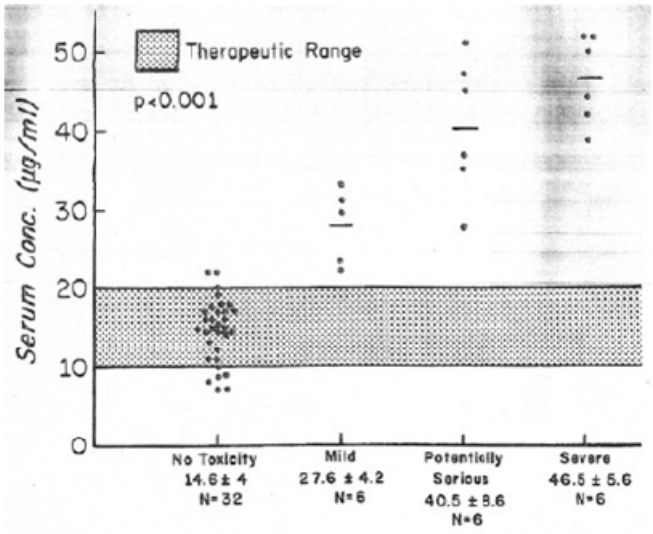

Figure 1.8. Mean theophylline serum concentrations versus the frequency and severity of toxic effects in 50 adult patients. Mild toxic events included nausea, vomiting, headache, and insomnia. Potentially serious toxic effects were limited to sinus tachycardia. Severe toxic side effects included life-threatening cardiac arrhythmias and seizures.

A number of studies have evaluated the effects of concomitant therapies on the pharmacologic outcome of theophylline, but they are specific to a single drug effect and sample population, and thus, have limited applicability. For example, Hoffman *et al*. investigated the effects of total body gamma irradiation on several drugs in rats.[99] They concluded that gamma radiation did not affect the theophylline dose necessary to induce seizures. Similarly, Hoffman *et al*. explored the potential for cyclosporine to potentiate the effects of theophylline in rats.[100] They observed that administration of cyclosporine reduced the theophylline concentration required to induce seizures and suggested that cyclosporine may increase the risk for generalized seizures during treatment with theophylline.

A survey of the literature reveals that the majority of PD studies involving theophylline focus on bronchodilation in asthmatic patients. One of the criteria used to diagnose asthma as well as monitor its condition is forced expiratory volume in one second (FEV_1), which is the amount of air that an individual can forcibly expire in one second. FEV_1 is typically measured in liters using a spirometer and is reported as a percentage of the amount of air that can be forcibly expired after full inspiration.

It has long been reported that the degree of bronchodilation (or percentage change in FEV_1) is linearly related to theophylline concentration when the concentration axis is log-transformed. Indeed, Mitenko *et al*. observed this relationship in six hospitalized asthmatic patients who were administered theophylline IV; no mathematical relationship defining the PD model was offered.[41] More recently, however, researchers have observed findings that challenge this correlation.

Flores-Murrieta *et al*. reported in 1999 that previous researchers had failed to model the PD outcomes of theophylline in asthmatic patients presenting with different degrees of airway obstruction.[101] The authors conducted a study in 15 asthmatic patients who were cleared of all confounding diseases. Patients were divided into two groups, the first consisting of 2 males and 6 females (mean age 40 ± 11 years) who displayed FEV_1 values of less than 50 % and the second comprised of 7 females (mean age 30 ± 12 years) with FEV_1 values between 50 and 70 % of ideal. A single 250 mg dose of theophylline was administered via IV infusion over 30 minutes and plasma drug concentration as well as FEV_1 was measured for a total of 12 hours. Plasma concentration-time data were modeled using a two-compartment open model consisting of a central and an effect compartment. A two-compartment model was determined to be optimal due to the delay between the appearance of theophylline in the plasma and bronchodilation. Plasma concentration in the effect compartment was correlated to FEV_1 via a sigmoidal E_{max} expression

$$E = E_0 + \left(\frac{E_{\max} \cdot C^{\gamma}}{EC_{50}{}^{\gamma} + C^{\gamma}} \right) \tag{1.18}$$

where E is the pharmacologic response, E_0 signifies the baseline response, E_{max} represents the theoretical maximum effect that can be attained, EC_{50} symbolizes the drug concentration at 50 % of the maximum effect, C is concentration at the effect site, and γ is the sigmoidicity constant, which dictates the steepness of the curve. The optimal fits for the response curves were determined using nonlinear regression. E_{max} pharmacodynamic models were previously demonstrated to be practical for correlating FEV_1 to theophylline concentration.[102]

The authors observed counterclockwise hysteresis loops in both groups when FEV_1 values were plotted against theophylline plasma concentration (Figure 1.9). This phenomenon suggests that the relationship between bronchodilation and theophylline concentration is indirect, possibly due to a delay in the equilibrium between plasma concentration and the site of drug action. Both groups were able to achieve comparable FEV_1 values when treated with theophylline; however, the EC_{50} value was higher for the severe patients, which increases their risk of experiencing a toxic event. The PK of the two groups were comparable. While these results confirm the indirect relationship observed by two other research groups,[103,104] they contradict the general viewpoint that the extent of bronchodilation is directly related to theophylline concentration. The authors concluded that by failing to segregate patients according to baseline airway function, earlier investigations may have overlooked the true exposure-response relationship for theophylline.[101]

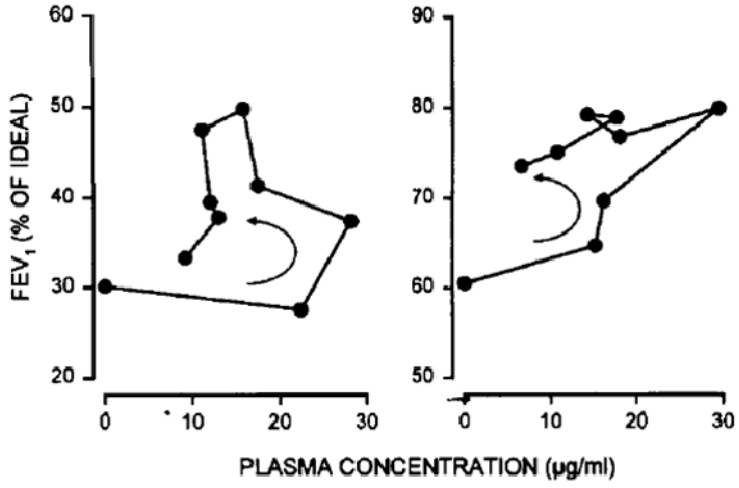

Figure 1.9. Plots of FEV_1 verses theophylline plasma concentration to patients with severe (left) and moderate (right) airway obstruction. The arrows indicate the counterclockwise hysteresis.

Despite years of research and a substantial number of literature publications, it is evident that the PD, much the same as the PK, of theophylline has yet to be fully elucidated. The author is unaware of any large-scale study assessing the safety and effectiveness of theophylline in a broad population encompassing the various physiologic, pathophysiologic, and other factors known to influence its action. As a result of these voids, theophylline remains a viable research candidate for PK and PD investigations.

1.3.7 Pharmacokinetic/Pharmacodynamic Simulation

The individual disciplines of PK and PD are merely two of the many discrete components situated on the continuum known as the drug discovery and development process. Integration of these components, however, convolves their contributions and offers knowledge that transcends each discipline to provide a more complete picture of the drug delivery system and its impact on patient well-being. Integrated PK and PD models bridge the relationship between dose and response. These models can then be used to better guide the development and ultimate utility of a drug delivery system. For example, integrated PK/PD models can be used to minimize the likelihood that a patient will experience an adverse event following commencement of a therapy. The same models can be used to understand how response changes between fed and fasted states or how the pharmacologic outcomes might be affected as a result of drug-drug interactions that alter the ADME of one or several of the concomitant drug therapies.[48]

Modern-day computers augment the arsenal of research tools available to scientists and clinicians. Technological advances in the areas of hardware and software have nearly eliminated computational deficiencies that were limiting factors several

decades ago. Thus, personal computers are powerful enough to conduct a variety of *in silico* simulations with varying underlying objectives. Numerical simulation can be expected to take a more prominent role in the drug discovery and development process as it compliments the FDA's Critical Path Initiative by utilizing innovative techniques to study the manufacturability, safety, and efficacy of candidate molecules and/or drug products.[9] Pharmacokinetics and pharmacodynamics are two pharmaceutical disciplines that utilize computer simulation extensively.

Once a PK model has been generated, a simulation platform can be constructed to assess the impact that random variation to the inputs has on the PK model output(s) (e.g., plasma concentration). The article by Gomeni *et al*. was previously introduced as it pertains to MCS.[70] Regarding the details of their PK modeling and simulation, the authors first implemented a convolution approach to estimate plasma concentration. Specifically, the *in vivo* delivery rate, which was modeled using the Weibull function, and the disposition and elimination time course, which was described by a two-compartment linear model, were convolved to predict plasma concentration. The authors then perturbed this model via MCS by randomly sampling values for the inputs (e.g., fraction of the dose absorbed, Weibull shape and time parameters) to better understand their influence on plasma concentration, and subsequently, receptor occupancy. Log-normal distributions for each input parameter were generated according to a predetermined coefficient of variation for that particular trial. The simulation evaluated plasma concentration for a given patient 24 hours after the 7th dose to better understand how variability of the inputs affected receptor occupancy.

Alternatively, Peters recently illustrated how physiologically-based PK simulation can be used to estimate PK model coefficients with two model drugs, verapamil and a proprietary compound no longer in development, in both rats and humans.[105] The primary objective was to demonstrate that physiologically-based PK simulation of plasma concentration-time profiles can reliably differentiate the underlying PK mechanisms, most notably the intestinal loss of orally administered drugs from first-pass hepatic metabolism, which are more often than not modeled in the same compartment as they are difficult to separate within clinical data. A generalized physiologically-based PK model comprised of a 9-compartment absorption model and a 14-compartment (i.e., 14-organ) somatic model was constructed to account for factors such the first-pass effect, intestinal loss, renal and biliary elimination, enterohepatic recirculation, and conversion of the metabolites back to the parent compound.

For a given drug, the first phase began by iteratively optimizing the estimates of clearance and the tissue partitioning coefficients of the 14 organs for IV administration using known physiochemical properties of the drug (i.e., permeability and solubility). Model performance was evaluated by assessing the goodness of fit for the predicted AUC with the mean of the actual AUC values. Once the predicted profile closely mirrored the actual IV plasma curve, the next step used the optimized clearance and distribution parameters to simulate the plasma concentration-time profile for oral administration. Assuming that solubility and permeability are the only two properties that determine the shape of the profile (i.e., that clearance and tissue distribution coefficients are not dependent upon the route of administration), differences between the actual and predicted concentration-time curves for oral administration should be attributable to intestinal loss

factors such as drug-induced gastric emptying delay, enterohepatic recirculation, gut wall metabolism, chemical degration, P-glycoprotein efflux, and/or variable absorption across the gut. These model parameters can then be modulated to reveal which factors are more than likely responsible for any differences between profiles. Peters emphasized that this method is functional even with inaccurate estimates of clearance and/or tissue partitioning coefficients. Further, she highlighted that physiologically-based PK modeling is a suitable approach for deconvolving absorption from distribution, metabolism, and excretion phenomena for orally administered drugs. The specific factors affecting the pharmacokinetics of the oral dosage form can then be differentiated through additional laboratory experimentation.[105]

It is evident that as physiologically-based modeling and simulation progresses, the concept of compartments is continually advancing from bodily systems or organs to cellular structures (e.g., enzymes, transporters). This presents the opportunity to model and simulate population variability on a cellular level, which will better elucidate how these disparities, which could be attributed to numerous factors (e.g., genetics, disease states, concomitant drug therapies), affect drug kinetics and action.[67]

A quantitative PK model is a prerequisite for predicting the clinical effectiveness of a given drug. Once such a model is available, it can be integrated with pharmacodynamic knowledge to study the safety and efficacy of treatment. Often times, simulation utilizes PK and PD models simultaneously to accelerate the drug development process. Such models are typically referred to as PK/PD models, which attempt to link the drug dosages to the clinical outcomes. The value of a PD model is dependent upon the merit of the corresponding PK model; a PD model that cannot accurately predict

pharmacologic response as a function of drug concentration has limited utility. Several case studies that investigated the potential impact of certain variables (e.g., dosing schedule, PK parameters, patient compliance) on the effectiveness of treatment were reviewed earlier.

Another important advantage of PK/PD simulation is its ability to help identify specific sub-populations that will benefit from treatment with the compound under investigation. As a rule of thumb, any compound that lacks efficacy will fail in clinical trials. Likewise, categorizing patients who will not respond safely or effectively to a given therapy is vital for the sponsor, the regulatory agencies, and the general public. With knowledge of a substance's ADME along with explicit inter-patient characteristics, simulation can elucidate specific patient conditions that preclude individuals from treatment. *In silico* investigations of patient variability in diverse populations has been coined *population pharmacodynamics*. As Michelson *et al*. indicates, responder populations can be identified using simulated patients generated from hypotheses or by fitting observed data to dose-response curves using any number of mathematical functions that account for covariates (e.g., age, gender, health factors).[106] Tools that identify responder populations should not only increase the number of available therapies, but should enhance the efficiency and effectiveness of clinical trials, which will ultimately hasten the time-to-market.

Eddy and Schlessinger published what is perhaps one of the most comprehensive examples of a simulation platform constructed to study a disease state.[107] The authors generated an extensive diabetes model and simulation platform referred to as the Archimedes model. While the Archimedes model transcends the classification of PK/PD

simulation, it is purposefully offered as a summary to the literature survey since it integrates many of the elements addressed thus far. The authors indicated that the model was too large to detail in a single manuscript, but the equations and corresponding assumptions are available in an online appendix (http://care.diabetesjournals.org). The Archimedes model was generated according to 3 basic criteria: (1) the model was to include all facets of the disease or its management that were considered valuable for investigating areas of interest, (2) the model was to be able to delineate clinically relevant features of the disease and its management, and (3) the level of detail incorporated was to be commensurate with its importance in the design of clinical trials. The Archimedes model was sequentially constructed and utilized in the following five stages: (1) develop a nonquantitative or conceptual description of the pertinent biology and pathology of diabetes (i.e., the variables and their relationships), (2) identify studies that focused on these variables and their underlying connection, (3) use the knowledge found in those investigations to link these variables via mathematical functions, (4) program these models into the simulation platform, and (5) perform numerical simulation using the platform. Individual models were tested and debugged during phase (4) of this series. Although the development criteria and phases were originally couched in terms of the Archimedes model, it should be noted that they are directly applicable to alternative simulation systems.

The Archimedes model was constructed using a system of differential equations and was coded using an object-oriented language known as Smalltalk. At the time of publication, it included numerous physiologic, pathologic, logistical, administrative, and economic factors including disease risk factors, incidence and progression of the disease,

glucose metabolism, symptoms, treatments, complications (e.g., coronary artery disease, congestive heart failure, deaths from diabetes and its associated complications, physiological dysfunctions as a result of disease (change in production of glucose as a result of an increase or decrease in insulin, etc.)), and differences amongst health care facilities (e.g., tests for pathophysiology of diabetes, re-admission rates). The structural relationship of the variables included in the diabetes model is summarized in Figure 1.10. The Archimedes model has the flexibility to assess three different treatment regimes (i.e., IV insulin, oral drugs, diet and exercise). Response to, for example, insulin is initially modeled by an individual's insulin factor from a distribution that characterizes the variation of the general population. This effect is then propagated through the various expressions that comprise the Archimedes model to account for confounding factors. The authors validated the Archimedes model in a subsequent article using actual clinical data from 18 trials. They concluded that the platform has the capacity to realistically model and simulate anatomic and pathophysiologic changes, treatments, and outcomes relevant to diabetes and its complications within the context of the available trial data.[108] They also indicated that the Archimedes model has the inherent flexibility to incorporate additional underlying knowledge regarding this disease state as it becomes available.[107]

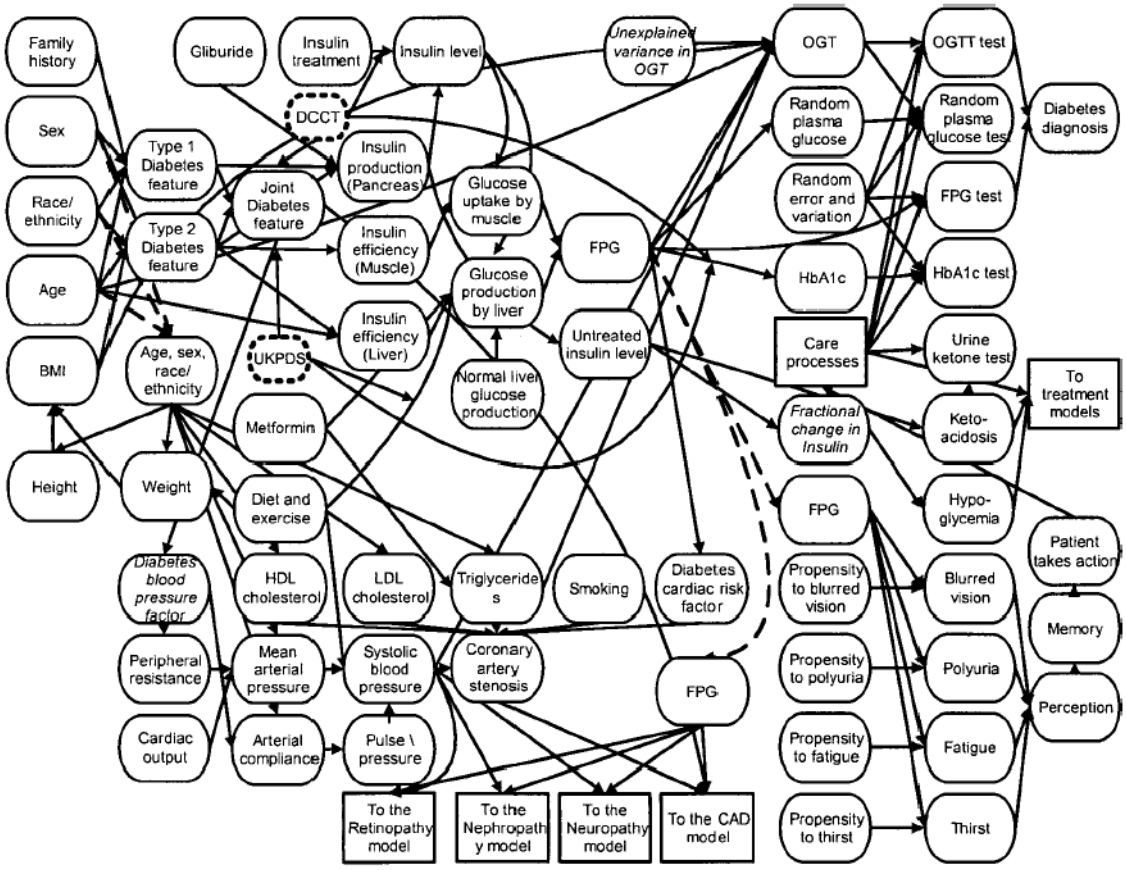

Figure 1.10. Schematic diagram detailing the structural relationship of the factors included in the Archimedes model of diabetes. Circles represent variables and lines indicate relationships. In general, the arrows directed into a variable represent one equation. Squares indicate other components of the model that are too complex to be shown here and have their own corresponding diagrams consisting of tens of variables and relationships. UKPDS and DCCT are acronyms for actual clinical trails and appear as circles with dashed boarders. These two trails were used to help construct the elements of the model illustrated in the figure. Reproduced from the work of Eddy and Schlessinger.

This work was founded on the central hypothesis that pharmaceutical process and product understanding can be simultaneously utilized to model the risk that final product quality imparts to clinical performance. The subsequent chapters, whole and in part, address one or more of the objectives stated in section 1.2. Chapter 2 describes the particular data, models, and assumptions used to construct the risk simulation platform, which is the medium used to redefine pharmaceutical quality in terms of risk by linking clinical attributes to production characteristics. Subsequently, the determination of the

conditional risk of product variation on clinical performance for the model drug delivery system is detailed. Chapter 3 builds upon the underlying relationships between product variation and clinical performance to generate design spaces conditioned on quantitative estimates of inefficacy and toxicity risk.

Multivariate data analysis and calibration are important elements of PAT and QbD as spectroscopic techniques are often used to acquire and/or enhance product and process knowledge. Chapter 4, therefore, demonstrates the utility of multivariate data analysis for elucidating the effects of various product and process variables on spectroscopic measurements. The influence of experimental design on spectroscopic variance is also considered. Chapters 5 and 6 described the role of net analyte signal theory and figures of merit in gauging the performance of calibration models, which will likely be integral components of future risk simulation efforts. Not only is it important to understand the performance of calibrations, it is critical to identify how their performance influences the prediction of factors that ultimately affect risk.

The penultimate chapter (7) seeks to unify the preceding topics via a hypothetical example revolving around the incorporation of PAT into a QbD production environment to ultimately control the clinical performance of the final product. Here, non-invasive spectroscopic techniques are strategically integrated prior to final product release to monitor those attributes which are potentially critical to quality. The corresponding role of process and control models in managing inefficacy and toxicity risk of the model drug delivery system is also addressed. Lastly, Chapter 8 provides a summary of the aforementioned work.

This dissertation explicitly links product characteristics to clinical performance using the proposed methodology. It is important to emphasize that these theophylline data were used to demonstrate one (of potentially several) approaches to directly relating product and patient characteristics. The clinical risk data were not, however, generated with the intent to suggest that theophylline regimens should or should not be altered.

Chapter 2: A New Definition of Pharmaceutical Quality: Assembly of a Risk Simulation Platform to Investigate the Impact of Manufacturing/Product Variability on Clinical Performance of a Model Theophylline Solid Oral Dosage System

2.1 Introduction

Around the turn of the 21st century, an informal retrospective survey of the pharmaceutical industry revealed that its progress (particularly with regard to manufacturing) was essentially stagnant and paled in comparison to other industrial sectors. Some individuals contend this was obvious and discussed within intellectual circles *ad nauseum*. Countless more, however, were reluctant to admit or failed to detect the issue at hand, and yet others who begrudgingly acknowledged the problem hurriedly placed blame on an overly restrictive regulatory system that penalized innovation. This was all set to change.

Taking initiative and holding themselves partially responsible, the FDA launched the CGMPs for the 21st Century campaign in 2002 to, in effect, "modernize" the pharmaceutical industry. The modernization commenced internally, and a new, risk-based regulatory architecture was created to refocus resources where they were needed most; areas that posed the greatest risk to the public. In turn, pharmaceutical companies were encouraged to adopt risk- and science-based approaches for drug discovery and development. Numerous initiatives, reports, and guidances followed (e.g., PAT, QbD), many of which promoted innovation and offered examples as to how the associated changes fit within the contemporary regulatory environment. These documents

underscore the need for collective change and emphasize several benefits that manufactures would reap from innovation.

"Quality" is explicitly or implicitly addressed in all of these documents. To date, the exact definition of quality in the pharmaceutical industry is unresolved, which is burdensome given that one of the primary objectives of the modernization initiative is to spur innovation to ultimately enhance pharmaceutical product quality. This uncertainty culminated in 2004 when Dr. Janet Woodcock, Director of the CDER at the FDA proposed re-defining pharmaceutical quality with regards to risk by linking production characteristics to clinical attributes.[19]

It is well understood that the clinical performance of any therapeutic regimen is dependent on a number of factors. For example, patient compliance dramatically influences safety and efficacy profiles. Consequently, researchers and clinicians invest time and energy to understand and control compliance rates. Manufacturing of the drug products also imposes a certain degree of risk on clinical performance. Despite its influence, little (if any) effort is devoted to quantifying the risk associated with manufacturing processes. If quality is to be re-defined in terms of risk, probabilistic relationships between production and clinical attributes must be established.

Cogdill and Drennen described an approach for relating manufacturing characteristics and clinical performance of a drug product.[20] They proposed the combination of probabilistic risk assessment (PRA) and Monte Carlo simulation (MCS) to relate elements such as raw material quality, product design, population statistics, dosing guidelines, and patient compliance estimates with pharmacokinetic (PK), pharmacodynamic (PD), and *in vitro-in vivo* correlation (IVIVC) models to remold

quality in terms of risk (Figure 2.1). The objective was to translate manufacturing and drug product attributes into probabilistic risk scores for toxicity and inefficacy. With these estimates, product and process design could then focus on minimizing risk to the patient.

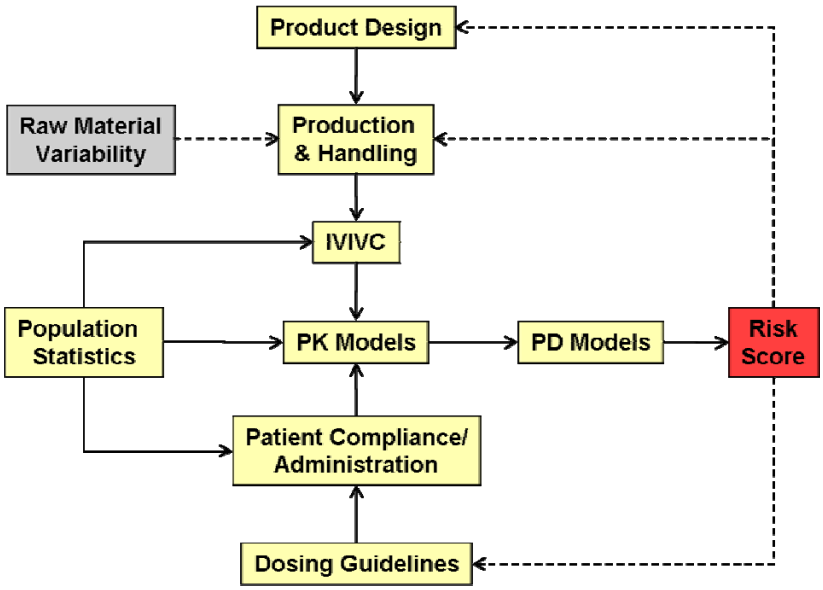

Figure 2.1. Schematic of the various model components that comprise the risk simulation platform. Figure adapted from Cogdill, RP; Drennen, JK. 2008. Risk-based quality by design (QbD): A Taguchi perspective on the assessment of product quality, and the quantitative linkage of drug product parameters and clinical performance. Journal of Pharmaceutical Innovation 3 (1): 23 – 29. Solid arrows represent components that are currently linked in the platform, whereas dotted arrows signify components/concepts that have yet to be incorporated.

This work used theophylline as a model drug to illustrate one potential method of relating manufacturing characteristics of a solid dosage system to clinical performance in simulated asthmatic patients displaying considerable inter-individual variability. The objectives were to (1) describe the structure for harnessing MCS and PRA to estimate risks of inefficacy and toxicity and (2) estimate the conditional risk of production characteristics on clinical performance for a model solid oral dosage system.

2.2 Materials and Methods

2.2.1 The Weibull Distribution

Waloddi Weibull formally introduced what is now referred to as the Weibull distribution in a 1939 monograph published by the Royal Swedish Institute for Engineering Research. As Arthur Hallinan, Jr. emphasized, there are (at least) five different mathematical formulae used to express the Weibull distribution, all which are equivalent following certain transformations.[109] It was his opinion that the array of formulae most likely created confusion, and, therefore, initial reservation concerning its applicability. Nonetheless, the Weibull distribution has been used to model numerous phenomena such as wind speed, failure (or reliability), and dissolution.[16,109,110]

Although originally hypothesized as a material function to describe the strength of materials subjected to stress, the Weibull distribution is now regarded as a flexible (i.e., generalized) statistical distribution. The distribution was proposed as a three-parameter distribution, characterized by a scale parameter (α), a shape parameter (β), and a location constant (c), but it is frequently utilized as a two-parameter ($c = 0$), and, at times, a one-parameter distribution (α and $c = 0$). It should be noted that the symbols used to represent these three parameters have varied over time, but these were intentionally selected so as to be consistent with the most recent literature. The following formulae will be presented according to the two-parameter functions; a location constant is employed to adjust the point at which there is a non-zero probability. Please refer to the article by Arthur Hallinan, Jr. for greater detail regarding the various three-parameter Weibull expressions.[109]

The cumulative distribution function (CDF), or the cumulative probability of occurrence for a given random variable (V) is described by the two-parameter Weibull function

$$f(V) = 1 - e^{-\left(\frac{V}{\alpha}\right)^{\beta}} \tag{2.1}$$

where V is the vector of points to be modeled ($v_i \geq 0$), β is the Weibull slope or shape parameter ($\beta > 0$), and α is the Weibull scale parameter ($\alpha > 0$). Similarly, the derivative of the CDF describes the probability density function (PDF), which is the probability distribution of a continuous random variable. The PDF is expressed by the equation

$$f'(V) = \frac{\beta}{\alpha} \cdot \left(\frac{V}{\alpha}\right)^{\beta-1} \cdot e^{-\left(\frac{V}{\alpha}\right)^{\beta}} \tag{2.2}$$

where V , β, and α are as previously defined. A random variable is said to be Weibull distributed if its CDF or PDF are adequately represented by Equations 2.1 or 2.2 (or the equivalent one or three-parameter functions). Several methods, including the Weibull graph technique, least-squares, and maximum likelihood, can be used to estimate the parameters.[109]

The Weibull distribution is considered flexible for a number of reasons. First, it is characterized by only 3 (or less) parameters, which is more straightforward than, for example, the five-parameter bi-variate normal distribution. Second, it mirrors the Rayleigh distribution and approximates the Gaussian distribution when the shape parameter is 2.0 and 3.6, respectively; these distributions can be classified as specific cases of the Weibull. Its ability to assume a range of values for the shape parameter, therefore, allows the distribution to more easily compensate for real-world variability. Finally, the Weibull distribution provides a reasonable fit to a variety of observed

distributions, which is evidenced by the assorted phenomena it has been used to model.[110] It was for these reasons that the Weibull distribution was selected to model certain data for this work.

2.2.2 Patient Simulation

Monte Carlo simulation has been shown to be an effective method for generating hypothetical patient populations in situations where it may be unreasonable or unethical to utilize humans.[64,111] For the work herein, MCS will be used to generate asthmatic patients ranging in age from 10 to 90 years; patients outside of this range were not modeled due to the lack of data pertaining to the targeted factors. The most significant factors affecting the disposition of theophylline, as determined by Jusko *et al*.,[46] will be specified to effectively represent inter-patient variability. All modeling and MCS simulations were performed using routines written in-house (Matlab, version 7.1, The MathWorks, Natick, MA ; PLS_Toolbox, version 3.0, Eigenvector Research, Inc., Manson, WA). Initially, data classifying the 2007 United States (US) population by age and gender were obtained from the US Census Bureau's International Data Base,[112] and statistics summarizing the prevalence of asthma within the US population during 2000 - 2005 were obtained from the Centers for Disease Control and Prevention (CDC).[113] Data for the prevalence of asthma specific to individuals older than 65 was not further delineated; thus, it was assumed that the prevalence (per 1000 subjects) monotonically decreased by 1.0 for each 10-year increment exceeding 70 years. A new distribution that approximated the asthmatic fraction of the total population within each age range was generated from the product of the US population and age-specific asthma rates.

An MCS routine, utilizing the exclusion method, assigned the age of each asthmatic patient based on the aforementioned relative distribution. The maximum number of allowable patients for each age range was determined by multiplying the number of desired patients by the relative distribution. To generate an initial estimate of age for a given patient, a single pseudo-random value was drawn from a uniform distribution generated on the unit scale (i.e., "rand.m" function within Matlab), was multiplied by the difference between the upper (i.e., 90) and lower (i.e., 10) age limits, and was added to the lower age limit. The estimate of age was rounded to the nearest integer. The "rand.m" function generates values between the closed interval of $[2^{-53}, 1-2^{-53}]$ and is theoretically capable of generating 2^{1492} values prior to repeating itself. If the estimate did not fall within the desired age range, the routine continued to generate estimations by resampling the uniform distribution until a satisfactory estimate was achieved. Likewise, if the estimate was within an age range where the maximum number of allowable patients had already been generated, the routine iterated until an acceptable age was attained. The distribution of age for the 100,000-patient population is summarized in Table 2.1. On average, approximately 375,000 iterations were required to assign the ages of the 100,000 patients.

Once the age of each patient was assigned, the gender of every patient was determined. Gender was resolved by drawing a number from a binomial distribution, which was generated using the "binornd.m" function in Matlab. The success probability (p) was set to the fraction of males in a specific age range.[112] A value of 1 signified a male, whereas 0 represented a female. The distribution of gender for the 100,000-patient population is summarized in Table 2.1.

Table 2.1. Age and gender distributions for the simulated patient population.

Age Range (years)	Males	Females	Combined Totals	Fraction of Total Population	Cumulative Fraction of Total Population	Fraction Male
0 - 4	0	0	0	0.000	0.000	0.000
5 - 9	0	0	0	0.000	0.000	0.000
10 - 14	5,324	5,226	10,550	0.106	0.106	0.505
15 - 19	5,305	4,927	10,232	0.102	0.208	0.518
20 - 24	4,348	4,247	8,595	0.086	0.294	0.506
25 - 29	3,781	3,692	7,473	0.075	0.369	0.506
30 - 34	3,251	3,224	6,475	0.065	0.433	0.502
35 - 39	3,583	3,410	6,993	0.070	0.503	0.512
40 - 44	3,728	3,846	7,574	0.076	0.579	0.492
45 - 49	4,117	4,246	8,363	0.084	0.663	0.492
50 - 54	3,957	4,088	8,045	0.080	0.743	0.492
55 - 59	3,443	3,714	7,157	0.072	0.815	0.481
60 - 64	2,634	2,971	5,605	0.056	0.871	0.470
65 - 69	1,901	2,153	4,054	0.041	0.911	0.469
70 - 74	1,420	1,643	3,063	0.031	0.942	0.464
75 - 79	1,034	1,415	2,449	0.024	0.966	0.422
80 +	1,187	2,185	3,372	0.034	1.000	0.352
0 - 80 +	49,013	50,987	100,000	1.000	1.000	0.490

Anthropometric reference data for the conditional distributions of total body weight (kg) and body mass index (BMI) of the US population for all ages during 1999 – 2002 were obtained from the National Health and Nutrition Examination Survey (NHANES), conducted by the National Center for Health Statistics (NCHS), CDC.[114] The weighted population means, standard errors of the means, and selected percentiles by sex, race, ethnic group, age, or age group were reported in this survey. Results, categorized by gender, are reported per year for individuals 1 to 19 and per decade for subjects 20 years and older. The 5^{th}, 10^{th}, 15^{th}, 25^{th}, 50^{th}, 75^{th}, 85^{th}, 90^{th}, and 95^{th} percentiles (i.e., the value in an ordered set of measurements for which x % of the observations lie below) were presented provided that the NCHS determined that the estimate was reliable. Unreliable figures were those that had a relative standard error (i.e., ratio of the standard error and the mean) greater than 30 %. The percentiles summarizing each age range for both weight and BMI were independently modeled using the two-parameter Weibull function (Equation 2.1). The weight and BMI values for males and females were modeled separately and the Weibull parameters were estimated using a least-squares approach. A shape and scale parameter were estimated for each age range modeled; the nominal percentiles were predicted using the reference percentile values[114] and the corresponding Weibull parameters to assess the goodness of fit. Irrespective of gender, the lowest coefficient of determination obtained for any of the individual models for weight or BMI was 0.951 and the median value across all models was 0.983. All anthropometric modeling was performed prior to the MCS for generating patients.

To assign weight and BMI metrics, the appropriate Weibull parameters were selected, first for weight and then for BMI, based on a patient's age and gender. The age-

and gender-categorized shape and scale parameters were used to generate a Weibull distribution (i.e., "wblrnd.m function in Matlab), from which a single value, representing the patient's weight or BMI, was extracted at random. Weight and BMI estimates were continually resampled until a value that fell within the restricted weight (20 – 130 kg) or BMI (10 – 50) range was obtained.

With the age, gender, weight, and BMI assigned, the remaining factors affecting the disposition of theophylline were specified. Statistics on cigarette smoking, marijuana use, alcohol consumption, and intake of oral contraceptives were obtained from the 2006 US Health survey conducted by the NCHS, CDC.[115] Statistics, reported as percent of total population, were taken for the latest recorded year. Data were (generally) reported for the ranges 12 – 13, 14 – 15, 16 – 17, 18 – 25, 26 – 34, and 35 years old and over. The last age bin was modified to include the expanded ranges 35 – 44, 45 – 64, and 65 – 90 years of age. Except where noted otherwise, the CDC data for greater than 35 years were represented identically in the three expanded age bins. For heavy alcohol drinkers, which was recoded as the percent of those who consumed alcohol, the values for 45 – 64 and 65 – 90 were adjusted downward (i.e., 8.8 % and 5.9 %, respectively) from the 10.4 % value reported for 35+ years based on the assumption that these individuals would pass away sooner than those who were not heavy drinkers. Additionally, the 3.1 % value reported for marijuana use for 35+ years was adjusted to 1.5 % for those individuals 65 – 90 years of age. For both marijuana and alcohol use, the percentage of female users was assumed to be 50 % of the observed rate for males; the values for heavy alcohol drinkers, however, were identical for males and females. Regarding the intake of oral contraceptives, the CDC data reported usage for minors in the range of 15 – 19 years. Thus, this value was

applied to the 16 – 17 age bin, while 0 % and 20 % were assumed for 12 – 13 and 14 – 15 years old, respectively. Additionally, a value of 0 % was assumed for the 45 – 64 and 65 – 90 ranges; oral contraceptive use was not recorded after 44 years. Regardless of age, a value of 0 % for oral contraceptive use was assumed for males. Data for cigarette smoking aligned with the amended age bins; therefore, they were used as reported.

The probability of suffering from congestive heart failure was modeled using data that described the prevalence of heart disease by age, which was obtained from the National Heart, Lung, and Blood Institute (NHLBI) Data Fact Sheet.[116] These data were reproduced from the 1976 – 1980 and 1988 - 1991 NHANES surveys; however, only data for 1988 – 1991 were used (Figure 5 of the NHLBI report). A data tracing program written in-house was used to estimate the prevalence (%) for the age-groups sampled. For individuals greater than 30 years of age, the probability of experiencing congestive heart failure (CHF) was approximated using the equation

$$\mathbf{p} \approx 3.92\,x10^{-3} \cdot (age)^2 - 0.22 \cdot (age) + 3.51 \qquad (2.3)$$

where **p** represents the probability of CHF and *age* is the age (in years) of the patient. Based on age- and gender-specific discharge frequencies for CHF cases recorded during 2004,[115] the resultant probability was multiplied by a factor of 0.75 if the patient was female. Additionally, all individuals 30 years of age or younger were automatically precluded from having congestive heart failure.

Sufficient gender- and age-specific data for the general use of barbiturates and benzodiazepines were unavailable. Therefore, inferences for the percent of the total population using each class of drug were made based on age and gender. These data are presented in Table 2.2.

Table 2.2. Gender- and age-specific data detailing the percent of the total simulated population using barbiturates and benzodiazepines.

	Age Range (years)	12 - 13	14 - 15	16 - 17	18 - 24	25 - 34	35 - 44	45 - 64	65 - 90
Barbiturate Use (% of total population)	Male	0.0	1.0	3.0	4.0	4.0	3.0	2.0	2.0
	Female	0.0	0.0	1.0	2.0	3.0	3.0	2.0	2.0
Benzodiazepine Use (% of total population)	Male	0.0	2.0	4.0	5.0	7.0	15.0	20.0	15.0
	Female	0.0	2.0	4.0	5.0	7.0	15.0	20.0	15.0

The likelihood of a patient presenting with any of the remaining factors shown to affect theophylline disposition (i.e., cigarette smoking, marijuana use, alcohol consumption, congestive heart failure, and intake of oral contraceptives, barbiturates, and benzodiazepines) was independently determined by randomly extracting a value from a binomial distribution, where p was set to the fraction of the total population that presented with a specific factor. Each factor was categorized based on age and gender, and therefore, p was selected based on the patient's age and sex. For congestive heart failure, p was merely the resultant probability estimated using Equation 2.3. An output of 1 from the binomial distribution represented a subject who displayed the given factor, whereas 0 indicated a patient who was negative for that factor.

Once all of the factors were accounted for, theophylline clearance was individualized for each patient according to the clearance cascade adapted from Jusko *et al*.[46] (Figure 2.2). It should be noted that the terminal nodes for marijuana use in individuals less than 40 years old who did not use oral contraceptives were excluded from the clearance cascade model. Data pertaining to the use of marijuana are subject to misrepresentation. Therefore, eliminating the second split based on marijuana use mitigated the uncertainty associated with accurately categorizing patients within the

cascade model. The terminal node on the clearance cascade was determined for each patient based on the individualized factors that predispose theophylline disposition. The percentage of the total 100,000-patient population that fell within each node is reported in Figure 2.2. Once it was determined which node best described a given patient, the mean and standard deviation of that particular node (Figure 2.2) were used to generate a normal distribution (i.e., "normrnd.m" function within Matlab), from which a single value, representing the patient's theophylline clearance, was extracted at random. Clearance estimates were restricted to 5 – 180 mL/hr/kg. If necessary, the distribution was resampled until the estimate was within the constrained range. The distribution of clearance for the 100,000-patient population based on the factors studied is summarized in Figure 2.3. Theophylline clearance was assumed to be constant throughout the course of treatment.

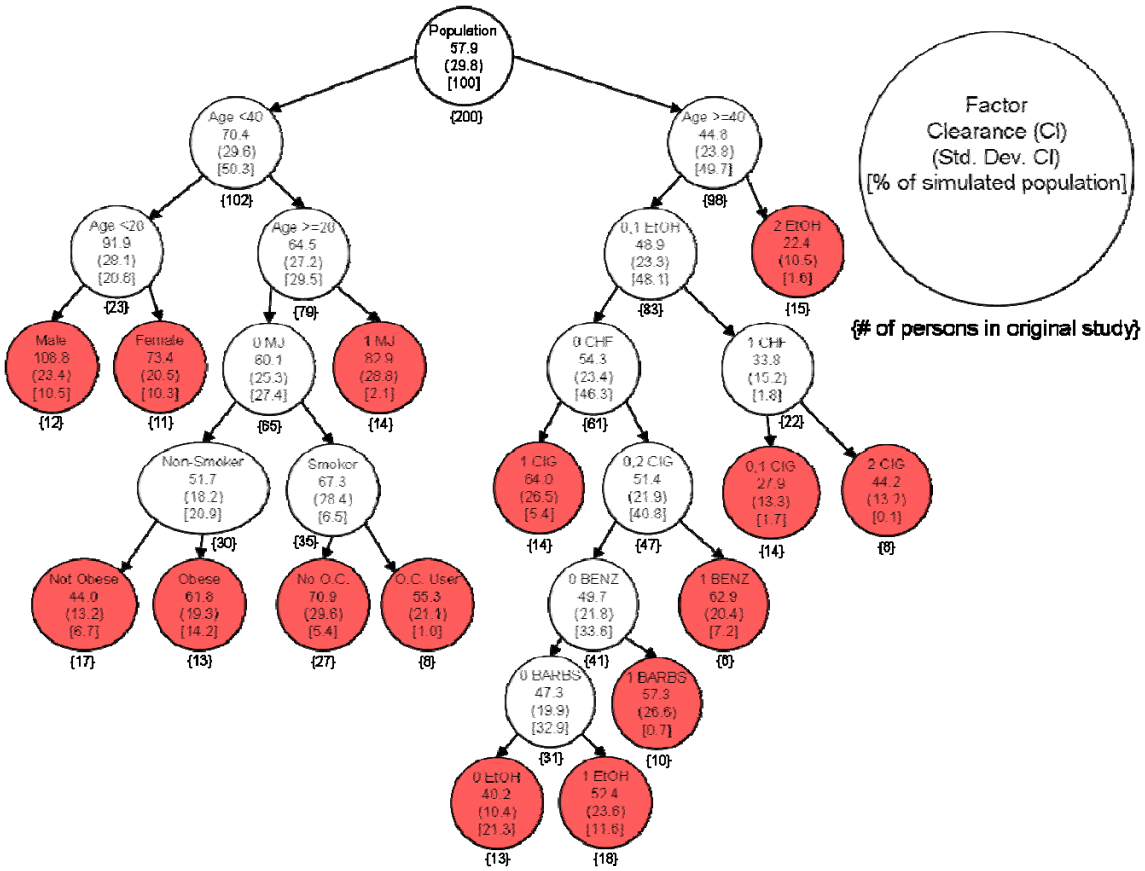

Figure 2.2. Clearance cascade detailing the average theophylline clearance for individuals classified according to numerous factors. Figure was adapted from Jusko, WJ, Gardner, MJ, Mangione, A, Schentag, JJ, Koup, JR, Vance, JW. 1979. Factors affecting theophylline clearances: Age, tobacco, marijuana, cirrhosis, congestive heart failure, obesity, oral contraceptives, benzodiazepines, barbiturates, and ethanol. Journal of Pharmaceutical Sciences 68 (11): 1358 – 1366. Both the number of individuals in the original study by Jusko et al. and the percentage of the 100,000 simulated population that fell within each node are indicated. All terminal nodes are shaded. 0, 1, and 2 signifies the extensiveness of a given factor as delineated in the original study. MJ = Marijuana; O.C. = Oral Contraceptive; EtOH = Alcohol; CHF = Congestive Heart Failure; CIG = Cigarette Smoker; BENZ = Benzodiazepines; BARBS = Barbiturates.

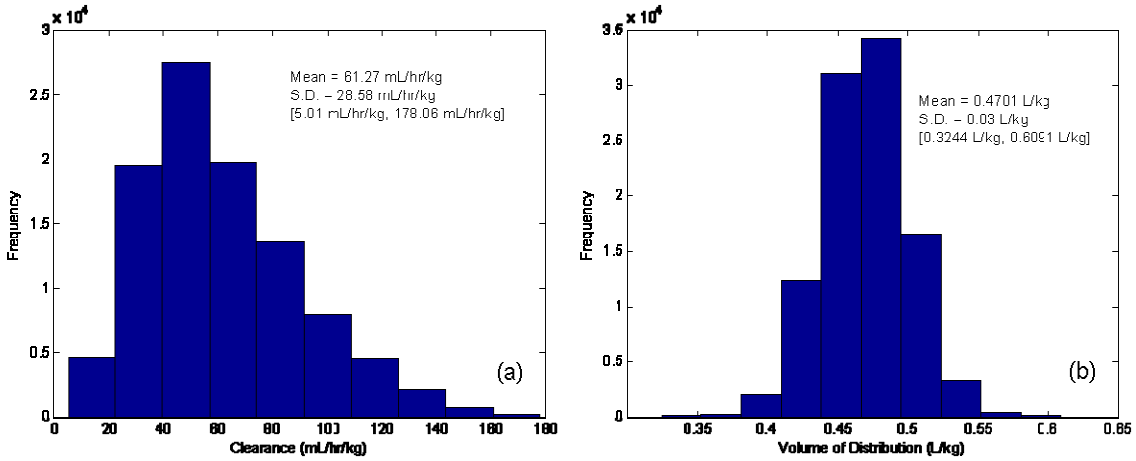

Figure 2.3. Frequency histograms of clearance (a) and volume of distribution (b) for the 100,000 simulated patients. The mean, standard deviation (S.D.), and range [,] of each parameter are also provided.

Finally, volume of distribution was assigned by randomly sampling a normal distribution defined by a mean volume of distribution of 0.47 L/kg and a standard deviation of 0.03 L/kg. While previous studies have assumed one (constant) average volume of distribution for all patients (e.g., 0.45 L/kg),[46,56] the author considered this to be more representative of the variability that would be encountered in actual patients. The volume of distribution values for the 100,000-patient population are summarized in Figure 2.3. Analogous to clearance, the volume of distribution was assumed to be constant for each patient during the course of simulated therapy.

2.2.3 Model Solid Oral Dosage Form

A solid oral theophylline dosage system that was previously formulated and processed at Duquesne University (Pittsburgh, PA) and compacted at a local pharmaceutical company was utilized for its estimations of manufacturing variability and clinical performance. The experimental details regarding these tablets have been described elsewhere.[117] Briefly, three separate manufacturing routes (i.e., direction

compression, roller compaction, and wet granulation) were used to produce 300 mg standard round bi-convex 3/8" diameter tablets on an 18-station high-speed rotary tablet press (model HT-AP1855-U/I, Elizabeth Hata). Eighteen distinct batches were manufactured using the direct compression and roller compaction routes, whereas 12 batches were produced via wet granulation. For the direct compression and roller compaction (Chilsonator, model IR 220, The Fitzpatrick Company) manufacturing methods, various combinations of anhydrous theophylline (BASF), lactose monohydrate (316 Fast Flo, Foremost Farms), microcrystalline cellulose (Avicel PH-102, FMC Biopolymer), and magnesium stearate (Spectrum Chemical) were processed and tableted. Tablets produced using the wet granules (planetary mixer, model 838F, Hobart) consisted of anhydrous theophylline, lactose monohydrate, magnesium stearate, and corn starch (Spectrum Chemical); a starch paste was used as the binding agent. For all three manufacturing methods, the compaction pressure was adjusted to yield target radial tensile strengths of 8, 11, or 14 kiloponds (kp). The nominal amount of theophylline was either 90 or 133 mg.

USP apparatus 2 (i.e., paddle) dissolution testing was performed using a Distek dissolution system (model 2100B) at a paddle speed of 50 revolutions per minute (RPM). The dissolution system was equipped with Hewlett-Packard UV-Vis spectrometer (model 8453) and a closed-loop automated sampler (Distek, Inc.). All dissolution testing was performed using deionized, de-aerated water as the medium in 900 mL Peak™ glass vessels at 37±0.1°C. The absorbance of theophylline was detected at 272 nm in 10 mm pathlength quartz flow cells following the construction of a standard curve. In total, 12 tablets per batch for each unique manufacturing route were assessed.

The Weibull function is often used to describe empirical dissolution data.[118] Dissolution profiles (i.e., percent theophylline released) of tablets produced via the direct compression, roller compaction, and wet granulation methods were modeled using the two-parameter Weibull function described by Equation 2.1 where V is the vector of dissolution time points. For dissolution modeling, the time constant (α) is often represented as $T_{63.2}$, the time at which 63.2 % of the drug is released. Once these data are fit to a Weibull distribution, the PDF (Equation 2.2) can be used to approximate the dissolution rate. Each dissolution curve was modeled by its reduction to a shape and a scale parameter. The distribution of dissolution shape parameters and dissolution time constants for the model system are presented in Figure 2.4. Two lines were fit to these data to represent the approximate maximum and median values for the dissolution shape parameter given the range of dissolution time constants (Figure 2.4).

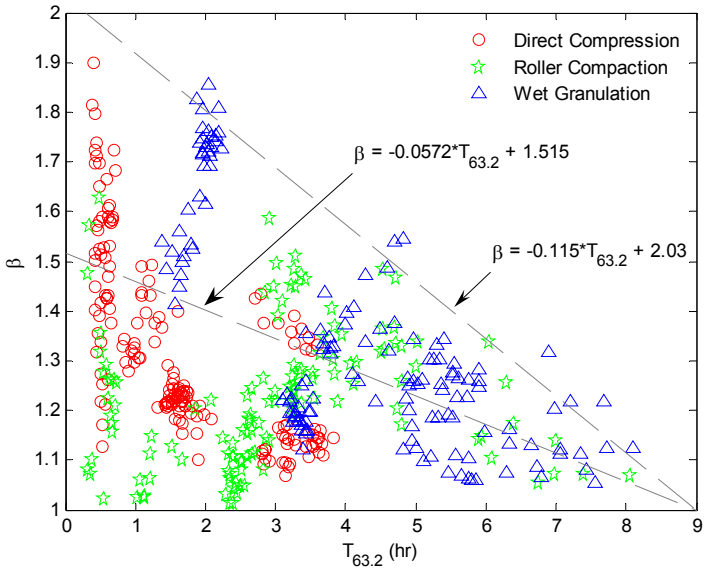

Figure 2.4. The distribution of dissolution shape parameters (ß) and dissolution time constants ($T_{63.2}$) for the model theophylline solid oral dosage system. Two lines were manually fit to these data to represent the approximate maximum and median values for the dissolution shape parameter given the range of dissolution time constants.

Content uniformity testing was also performed on dissolved tablets using a UV-Vis spectrometer (Hewlett-Packard, model 8453). Tablets were pulverized and dissolved in deionized water. The absorbance of theophylline was detected at 272 nm in 10 mm pathlength cuvettes at 25 °C using a standard curve independent of the one implemented for dissolution testing. Uniformity of tablets produced from all three manufacturing routes was assessed. In total, 10 tablets per batch for each unique manufacturing routine were analyzed.

Dissolution time constants and content uniformity estimates were segregated by batch to generate estimates of manufacturing variability. Intra-batch refers to the standard deviation of mean-centered observations within a batch, whereas inter-batch denotes the standard deviation of the mean observations across all batches (e.g., content uniformity, dissolution time constant).

2.2.4 IVIVC Model

Hussein and Friedman modeled the release and absorption characteristics of several novel, self-prepared, sustained-release (SR) theophylline formulations in addition to two commercial SR products (i.e., Theotrim and Theo-Dur).[92] Specific details regarding the materials and methods can be obtained from their original publication. Briefly, USP apparatus 1 (i.e., rotating basket) dissolution testing was performed at 100 RPM in 600 mL vessels. The first 2 hours of testing was conducted in 400 mL of simulated gastric fluid containing pepsin, after which, the medium was replaced with 400 mL of simulated intestinal fluid containing pancreatin and was monitored for an additional 10 hours. Theophylline concentration was subsequently determined by HPLC analysis. Six healthy volunteers were administered each formulation in a crossover study

observing a washout period of 3 weeks. Theophylline plasma concentrations were estimated using an HPLC method and the percentage of the theophylline dose absorbed was then determined using the Wagner-Nelson method.

The mean (n = 4) *in vitro* dissolution percent released and the mean (n = 6) *in vivo* percent absorbed profiles obtained from the volunteers for the said dosage forms were used to construct a Level A IVIVC model via a deconvolution approach (refer to section 1.3.4). These data correspond to Figures 1 and 3 of their original publication.[92] Only the values for formulations T-1, T-1-A, T-2-A, and Theotrim were modeled. A graph tracing program was used to extract quantitative data from the figures describing the percent released and percent absorbed at the time points sampled. Subsequently, these data were modeled using a two-parameter Weibull function (Equation 2.1); the fitted shape and scale parameters were used to estimate the instantaneous rates of release and absorption for the *in vitro* and *in vivo* data, respectively (Equation 2.2). Finally, the instantaneous dissolution rates were fitted using the Power Law to determine the IVIVC function (Figure 2.5). The resultant nonlinear function for transforming *in vitro* release to *in vivo* absorption was determined to be

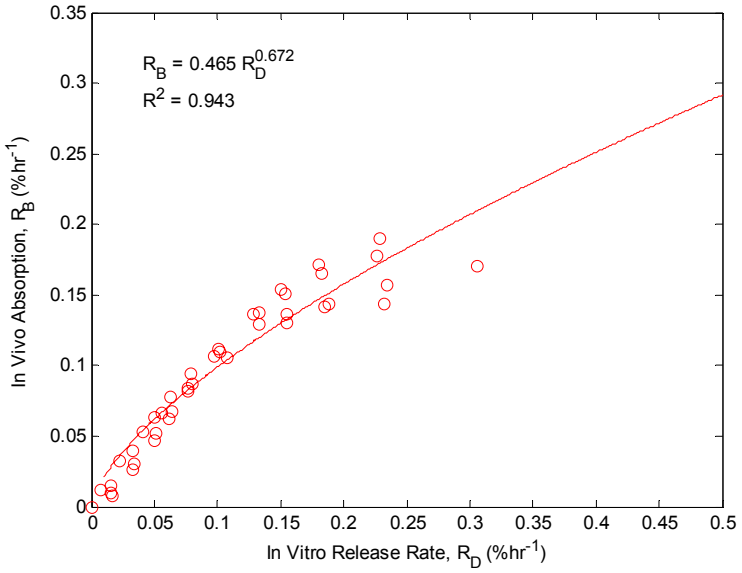

Figure 2.5. Plot of the in vivo absorption rate versus the in vitro dissolution release rate. The IVIVC model was fit using a Power Law function.

$$R_B = 0.465 \cdot R_D^{\ 0.672} \tag{2.4}$$

where R_B is the *in vivo* absorption rate, R_D is the *in vitro* dissolution release rate, and 0.465 and 0.672 are the scale factor (unitless) and Power Law parameter (unitless), respectively; the coefficient of determination for this function was 0.943.

2.2.5 PK Model

First-order pharmacokinetics by means of a one-compartment open model were assumed to adequately describe theophylline plasma concentrations following administration of the solid oral dosage form. Since multiple dosages were administered throughout the course of therapy, the principle of superposition was applied.[82] Superposition provides the opportunity to forecast plasma concentration-time curves based on the viewpoint that drug levels from successive doses are linearly additive. The

superposition principle assumes that the pharmacokinetics of the drug are not dose-dependent and that the drug is eliminated by first-order kinetics, which are reasonable assumptions for the administration of theophylline (refer to section 1.3.5.2). The change in theophylline plasma concentration as a function of time was modeled using the equation

$$\frac{dC_p}{dt} = \frac{D \cdot S \cdot A}{V_d \cdot W}\left[\beta \cdot \alpha^{-\beta} \cdot t^{\beta-1} \cdot e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\right]^{P} - \frac{Cl}{1000 \cdot V_d} \cdot C_p \qquad (2.5)$$

where C_p is the theophylline plasma concentration (mg/L), t is the time (hr), S is the optional scaling factor (unitless), D is the dose (mg), V_d is the volume of distribution (L/kg), W is the patient's total body weight (kg), β is the Weibull shape parameter (unitless), α is the Weibull time constant (hours), A is the IVIVC scale factor (unitless), P is the IVIVC Power Law parameter (unitless), and Cl is clearance (ml/kg/hr). All simulations were performed with S at a constant value of 1.0. This equation was derived to characterize the change in theophylline plasma concentration based on the relationship between the dissolution and absorptions rates and the individualized patient pharmacokinetic parameters. The output is mg/L/hr of theophylline.

Similar to the work of Buchwald,[83] theophylline input was modeled using sigmoidal lag time and cut-off coefficients where absorption was assumed to be 100 % of the maximum rate after 0.5 hours (i.e., lag time) and the absorption potential was reduced to 50 % after 8 hours (i.e., cut-off) to simulated time-dependent phenomena. These

coefficients were used to adjust the input (I) of theophylline through the following series of equations

$$I = \frac{D \cdot S \cdot A}{V_d \cdot W}\left[\beta \cdot \alpha^{-\beta} \cdot t^{\beta-1} \cdot e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\right]^{P} \tag{2.5a}$$

$$lagtime = \frac{1}{1+e^{-(t-0.5)\left(\frac{15}{0.5}\right)}} \tag{2.5b}$$

$$cutoff = 1 - \frac{1}{1+e^{-(t-8)}} \tag{2.5c}$$

$$\frac{dC_p}{dt} = \left(I \cdot lagtime \cdot cutoff\right) - \frac{Cl}{1000 \cdot V_d} \cdot C_p \tag{2.5d}$$

Figure 2.6 shows the lag time and cut-off coefficients for one 12-hour dose. Equation 2.5d is analogous to the original PK model with the exception of the lag time and cut-off terms. The numerical solution to Equation 2.5d was obtained via a Matlab-based differential equation solver (i.e., ode23).

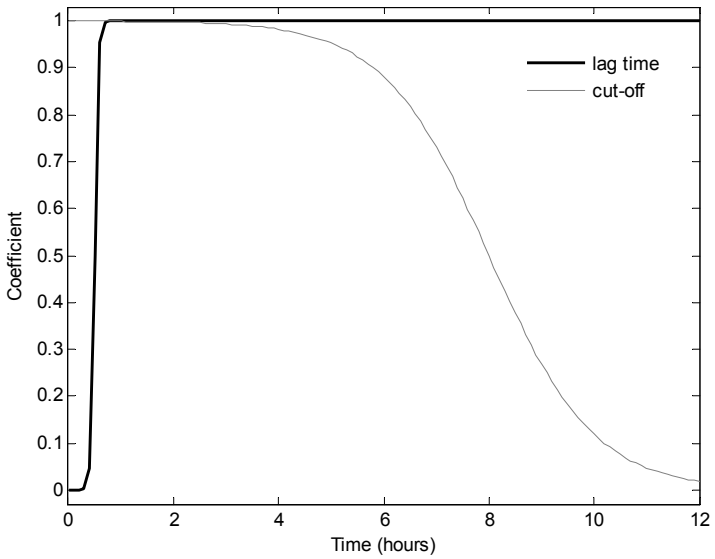

Figure 2.6. Plot illustrating the lag time and cut-off absorption coefficients over a 12-hour window.

2.2.6 PD Model

One or more probabilistic PD models are most desirable for a risk assessment tool such as the one described herein. For example, access to several PD models that characterize the probability of efficacy, the probability of multiple adverse events (e.g., headache, vomiting, seizure) and the covariance between these observations is optimal. More often than not, said models will not be available during the initial stages of risk assessment. Furthermore, efficacy may be characterized by various responses (e.g., forced expiratory volume, number of asthmatic attacks, quality of life), further obscuring the dose-response relationship and thus, the probability of a given outcome. Therefore, it is necessary to assume an underlying model, which can be replaced, augmented, or combined with additional models as the level of understanding increases.

A probabilistic-based PD model detailing the general efficacy and toxicity of theophylline was not readily available. Therefore, the authors chose to implement a model for a hypothetical drug, which also had a therapeutic range of 10 – 20 mg/L (labeled Figure 1.7 in the reference).[93] Data points were reproduced using the tracing program and were fitted using a sigmoid function. The PD model originally described the probability (%) of efficacy and toxicity as a function of drug concentration. The estimated sigmoid functions for efficacy and toxicity are provided in Equations 2.6a and 2.6b, respectively

$$\hat{P}_E = \frac{74.77}{1 + e^{-[(Z - 0.96)/9.70]}} + 3.83 \qquad (2.6a)$$

$$\hat{P}_T = \frac{74.29}{1 + e^{-[(Z - 1.40)/19.40]}} + 3.77 \qquad (2.6b)$$

where $\hat{P}_E$ and $\hat{P}_T$ are the predicted probabilities (%) for efficacy and toxicity, respectively, and Z is the vector of log-transformed theophylline plasma concentrations (mg/L). The PD model was adapted to describe the probability (%) of inefficacy and toxicity as a function of theophylline plasma concentration (Figure 2.7). Inefficacy estimates were generated by subtracting the efficacy probabilities from 100 %. No specific distinctions were made between various inefficacious or toxic events; the risk of observing, for example, a headache or a seizure was identically weighted.

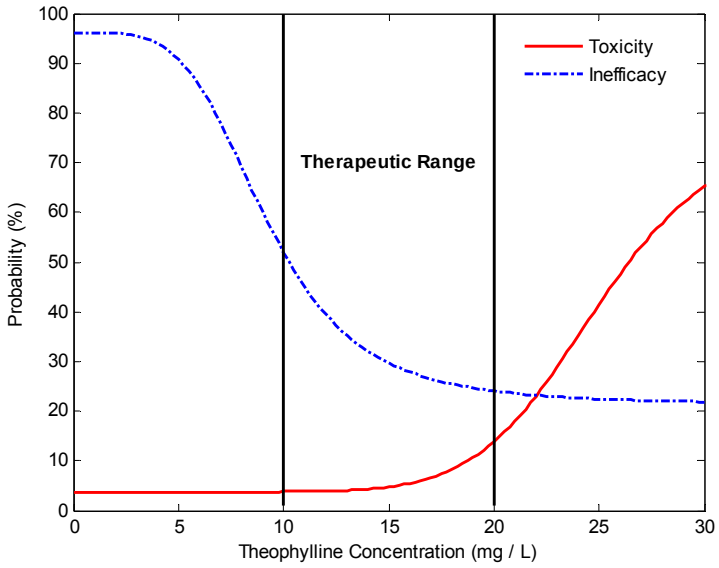

Figure 2.7. Pharmacodynamic model for theophylline describing the probability of observing a toxic or an inefficacious event as a function of theophylline plasma concentration. Figure adapted from DiPiro JT, Spruill WJ, Blouin RA, Pruemer JM. 2002. Concepts in Clinical Pharmacokinetics 3rd ed., New York: American Society of Health-System Pharmacists, Inc. (ASHP). p 279.

2.2.7 Dosing

Each patient was subjected to an iterative dosing scheme where his/her initial dose (D) was estimated using the equation

$$D = \frac{C_T \cdot Cl \cdot \frac{1}{1000} \cdot W \cdot Q}{F} \tag{2.7}$$

where C_T is the target plasma concentration (i.e., the median concentration of the therapeutic window, defined as 10 – 20 mg/L), Cl is the individual's theophylline clearance as predicted by the Jusko *et al*. model (mL/hr/kg),[46] W is the patient's weight (kg), Q is the time interval between doses (i.e., 12 hr), and F is the fraction of dose absorbed systemically (unitless). A constant value of 0.8 was assumed for F, which is comparable to values reported for other oral theophylline formulations.[32] Based on the nominal amount of theophylline assumed to be in each tablet (i.e., 100 mg), the number of tablets necessary to yield the initial dose was estimated (the number of units was rounded to the nearest integer). Following a period of time assumed to be sufficient to reach steady-state (i.e., 5 doses), the patient's plasma concentration was estimated via the PK model (Equation 2.5). If the dose was found to be inadequate, it was incrementally adjusted (either increased or decreased depending on whether it was too low or too high, respectively) until the iterative dosing scheme converged on a satisfactory dosage. If, however, the dose was adequate to yield a plasma concentration between the minimum effective concentration (MEC) and the minimum toxic concentration (MTC), treatment was initiated and the patient was administered the said dose for the duration of the trial period. On average, approximately 1.2 dose adjustments per patient were necessary for a given sub-population of 1500 individuals (data from dose adjustment iterations are not included in calculation of risk scores).

2.2.8 Risk Simulation

All risk simulations employed a MCS routine independent of that used to generate the patients. The simulation platform was constructed such that the user is able to specify the age range of the population to be tested (recall that age is a covariate for all other patient factors), as well as the manner in which the simulation terminates. For this option, the user can specify the number of individuals to be included in the sample population. Otherwise, the simulation can be set to iterate until specific convergence criteria are reached. All simulations presented in this work were terminated using convergence criteria. The user is also able to specify patient compliance and manufacturing variability estimates, as well as details concerning the drug and its corresponding therapy. These were determined or assumed for the model theophylline solid oral dosage system tested herein (Table 2.3).

Table 2.3. Summary of the manufacturing variability metrics and the treatment parameters used during simulation.

Manufacturing Metrics	
Intra-Batch RSD of Dissolution Time Constant	0.06
Inter-Batch RSD of Dissolution Time Constant	0.03
RSD of Intra-Batch Content Uniformity	0.03
RSD of Inter-Batch Content Uniformity	0.01
Simulation Parameters	
Length of Therapy (days)	30
Time Interval between Doses (hours)	12
Standard Deviation of Dosing Interval (hours)	1
Therapeutic Window (mg/L)	[10 - 20]
Rate of Compliance (% of doses taken)	90
Fraction of Dose Absorbed (unitless)	0.8
Dissolution Time Constant (hours)	5.0
Nominal Theophylline Amount (mg)	100.0

Given that one of the principal objectives of this work was to estimate the conditional risk of product quality variation on clinical performance, the simulator was

assembled such that the user could allow or prohibit the estimates of certain factors to be sampled according to their underlying distributions. These factors included the inter- and intra-batch relative standard deviation (RSD) of the dissolution time constants, the inter- and intra-batch RSD of content uniformity, the rate of patient compliance, and the standard deviation of the dosing interval. Thus, the risk simulation platform user was required to set variability "flags" prior to the start of a simulation that turned the factors "on" or "off" to assess their effect on risk. If a factor was turned off, its estimate was consecutively set to the same value, whereas if it was allowed to vary, the estimate was influenced by the level of variability (some would refer to this as "quality") or rate of adherence. For example, the scenario where each tablet contains the same amount of active (as per the label claim) represents the highest degree of quality (minimal variability) in terms of content uniformity.

A total of 4 variability flags were to be set by the user: inter- and intra-batch dosage variability, patient compliance variability, and dosing variability. All, none, or a combination of these variability flags could have been turned on during the course of a given simulation. When the dosing time interval was subject to variation, each dosing time was altered by the addition of a pseudo-random number drawn from a normal distribution with zero mean and unit standard deviation (i.e., "randn.m" function within Matlab); the random number was multiplied by the standard deviation of the dosing interval (Table 2.3) before it was added to the particular dosing time. Otherwise, doses were administered at their scheduled times. For simulations where patient compliance was variable, compliance was modeled using a binomial distribution where the success probability was set to the assumed patient compliance (% of doses taken); a value of 0

denoted a missed dose. Patients were prohibited from missing two consecutive doses. Noncompliance was prohibited during the patient-specific iterative dosing schedule. Otherwise, all doses were assumed to be taken.

The remaining two variability flags pertain to the dosage form itself. For instances where inter-batch variability was initiated, the initial dose administered to a patient (D') was randomly selected from a normal distribution, whose mean was set to D (Equation 2.7), and whose standard deviation was set to the inter-batch RSD of content uniformity (Table 2.3) multiplied by D. This estimate then remained the mean nominal dosage for that patient throughout the course of treatment (e.g., 30 days). Additionally, the inter-batch variability flag also altered the dissolution time constant; α' was randomly selected from a normal distribution whose mean was the nominal time constant (α) and whose standard deviation was set as the inter-batch RSD of the dissolution time constant multiplied the nominal α (Table 2.3). Again, α' was held constant for the duration of therapy. When intra-batch variability was prompted, each dose administered to a given patient was adjusted from the nominal amount (either D or D', depending on whether or not inter-batch variability was triggered) to reflect the level of variability around the mean for the current batch. This was accomplished by randomly selecting the current dose from a normal distribution of mean D or D' and standard deviation of D or D' multiplied by the intra-batch RSD of content uniformity (Table 2.3). Additionally, the intra-batch variability flag also altered the dissolution time constant for each dose; it was randomly selected from a normal distribution whose mean was the nominal time constant (α or α', depending on whether or not inter-batch variability was triggered) and whose standard deviation was set as the intra-batch RSD of the dissolution time constant

multiplied by α or α'. Otherwise, the dose estimated using Equation 2.7 was successively administered assuming a constant dissolution time constant when dosage form variability was not assessed. The simulation assumed that the one-month drug supply (for each patient) was drawn from a single batch.

The final parameter that needed to be addressed was β. The Weibull shape parameter was estimated for each dose using the linear relationships describing the approximate median and maximum values of β as a function of α (Figure 2.4). Specifically, β was randomly selected from a normal distribution. The mean of the normal distribution was set as the median value of β and the standard deviation was set to the standard deviation of β, which was estimated using the 99.9 % confidence interval for a normal distribution and the difference between the maximum and median shape parameters for a given time constant. The minimum allowable value for β was 1.01.

With all of the parameters set, the program commenced by first excluding those patients not meeting the age criteria, that is if the criteria differed from 10 – 90 years. Each patient was randomly selected from the sub-population and dosed accordingly. Once the appropriate dose was determined for each patient, he/she was administered treatment. Throughout the course of the therapy, a patient's theophylline plasma concentration was monitored by integrating Equation 2.5. Plasma concentrations were estimated 6 times per hour. These data were stored and superimposed over the course of treatment. A frequency histogram summarizing theophylline plasma levels was generated for each patient; responses were segregated (i.e., binned) into 0.25 mg/L intervals.

Probabilistic estimates of observing inefficacious and toxic events were predetermined for theophylline concentrations ranging from 0 to 100 mg/L at 0.25 mg/L increments using the PD sigmoid functions (Equations 2.6a and 2.6b). Using these concentration-based likelihoods, risk estimates (or scores) were generated after each patient was treated. First, the plasma concentration histograms were aggregated (i.e., data within each concentration bin were amassed for all patients tested). Next, the aggregated plasma concentration data was transformed into a CDF. Provided that both the PD functions and the pooled CDF were generated using the same concentration axis, plots of the inefficacy risk scores versus the aggregated CDF data and the toxicity risk scores versus the aggregated CDF data were generated. These plots were used to interpret the percentage of the population that had a risk score at or below a given value (i.e., the likelihood of observing an adverse event within a sample population given the observed plasma concentrations). Example plots for inefficacy and toxicity are shown in Figure 2.8a and Figure 2.8b, respectively. These plots illustrate that 95 % of the sample population had an inefficacy risk score less than or equal to 25.62 % and a toxicity risk score of less than or equal to 8.01 % for the given trial simulation. In other words, 95 % of the population was treated such that there was a maximum likelihood of 25.62 % and 8.01 % for observing an inefficacious or toxic event, respectively.

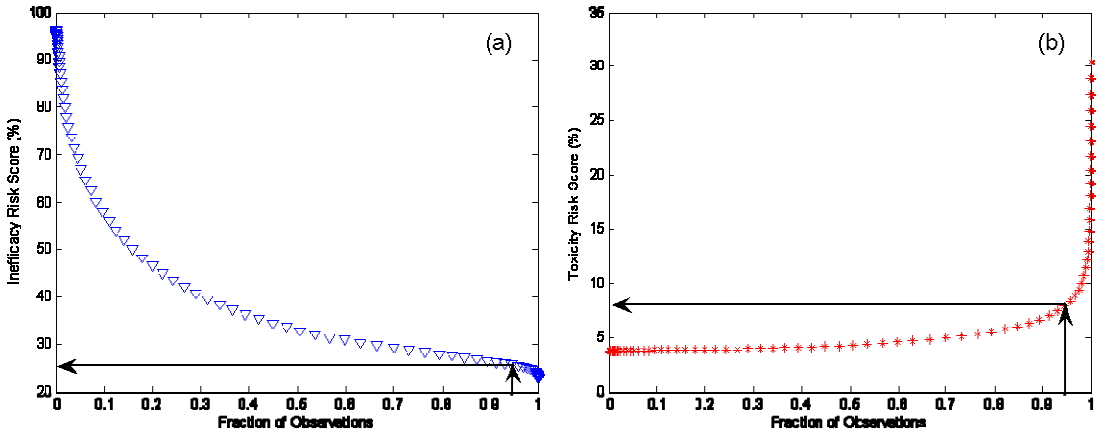

Figure 2.8. Plots of inefficacy (a) and toxicity (b) risk scores versus the fraction of observations for the sample population tested. These data were interpolated (solid lines) to determine the percentage of the sample population (95 %) that had a risk score less than or equal to a given value. The sample population was treated such that there was a maximum likelihood of 25.62 % and 8.01 % for observing an inefficacious or toxic event, respectively.

Rather than reporting multiple risk scores for both inefficacy and toxicity, it was desirable to summarize the risk to a sample population with a single risk score for each adverse event. Thus, the empirical CDF/PD function plots were interpolated to yield a single risk score corresponding to a CDF probability of 0.95 for both inefficacy and toxicity. A risk score summarizing those tested was generated for each addition of a patient. The number of iterations conducted was not fixed; rather, the risk simulator continued to test additional patients until the risk scores for inefficacy and toxicity both stabilized below a certain oscillation threshold. Stability of risk assessments was assessed by calculating the absolute fractional change of the median risk score (Δ) observed by adding one additional patient to the sample population using the equation

$$\Delta = \frac{abs\left[median\left(\overset{n}{\underset{i=1}{RS_i}}\right) - median\left(\overset{n-1}{\underset{i=1}{RS_i}}\right)\right]}{median\left(\overset{n}{\underset{i=1}{RS_i}}\right)} \tag{2.8}$$

where *RS* indicates the risk score for the i^{th} observation and *n* represents the number of patients assessed. Patients were consecutively tested until the variability of the risk estimates for both inefficacy and toxicity were below the difference threshold of 10^{-4}. Furthermore, the absolute change was required to retain a value below the threshold for 250 consecutive patients before the simulator converged on the risk estimates; these criteria were required for both inefficacy and toxicity. Therefore, two risk scores, one for inefficacy and one for toxicity, were generated for each trial simulation.

2.2.9 Experimental Design

A 2x2x2x2x2x2 full factorial design was generated in Matlab using the "fullfact.m" function to assess the effects of manufacturing variability and patient compliance on clinical performance (Table 2.4). Two levels for each factor were tested, which corresponded to the presence or absence of variability (i.e., factor on or off, respectively). The six factors assessed were the inter- and intra-batch RSD of the dissolution time constants, the inter- and intra-batch RSD of content uniformity, the rate of patient compliance, and the standard deviation of the dosing interval. A value of 1 signified the presence of variability, whereas 0 represented its absence. Each row in the design represents an independent risk simulation trail. The full factorial experimental design was performed in triplicate, which required a total of 192 simulations. The simulation run order for each replicate of the design matrix was randomized. The age range for patient inclusion was not altered from that of the general population.

Table 2.4. Summary of the 2x2x2x2x2x2 full factorial experimental design. A value of 1 signifies the presence of variability, whereas 0 represents its absence.

Trial Number	Intra-Batch β	Inter-Batch β	Intra-Batch Content Uniformity	Inter-Batch Content Uniformity	Patient Compliance	Standard Deviation of Dosing Time Interval
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	1	1	0	0	0	0
5	0	0	1	0	0	0
6	1	0	1	0	0	0
7	0	1	1	0	0	0
8	1	1	1	0	0	0
9	0	0	0	1	0	0
10	1	0	0	1	0	0
11	0	1	0	1	0	0
12	1	1	0	1	0	0
13	0	0	1	1	0	0
14	1	0	1	1	0	0
15	0	1	1	1	0	0
16	1	1	1	1	0	0
17	0	0	0	0	1	0
18	1	0	0	0	1	0
19	0	1	0	0	1	0
20	1	1	0	0	1	0
21	0	0	1	0	1	0
22	1	0	1	0	1	0
23	0	1	1	0	1	0
24	1	1	1	0	1	0
25	0	0	0	1	1	0
26	1	0	0	1	1	0
27	0	1	0	1	1	0
28	1	1	0	1	1	0
29	0	0	1	1	1	0
30	1	0	1	1	1	0
31	0	1	1	1	1	0
32	1	1	1	1	1	0
33	0	0	0	0	0	1
34	1	0	0	0	0	1
35	0	1	0	0	0	1
36	1	1	0	0	0	1
37	0	0	1	0	0	1
38	1	0	1	0	0	1
39	0	1	1	0	0	1
40	1	1	1	0	0	1
41	0	0	0	1	0	1

Trial Number	Intra-Batch β	Inter-Batch β	Intra-Batch Content Uniformity	Inter-Batch Content Uniformity	Patient Compliance	Standard Deviation of Dosing Time Interval
42	1	0	0	1	0	1
43	0	1	0	1	0	1
44	1	1	0	1	0	1
45	0	0	1	1	0	1
46	1	0	1	1	0	1
47	0	1	1	1	0	1
48	1	1	1	1	0	1
49	0	0	0	0	1	1
50	1	0	0	0	1	1
51	0	1	0	0	1	1
52	1	1	0	0	1	1
53	0	0	1	0	1	1
54	1	0	1	0	1	1
55	0	1	1	0	1	1
56	1	1	1	0	1	1
57	0	0	0	1	1	1
58	1	0	0	1	1	1
59	0	1	0	1	1	1
60	1	1	0	1	1	1
61	0	0	1	1	1	1
62	1	0	1	1	1	1
63	0	1	1	1	1	1
64	1	1	1	1	1	1

A 2x2x2x2 full factorial design was also generated to assess the effects of manufacturing variability excluding all influences from patient compliance and dosing time variability (Table 2.5). Again, two levels for each factor were tested where a value of 1 signified the presence of variability and 0 represented its absence. The four factors assessed were the inter- and intra-batch RSD of the dissolution time constants, and the inter- and intra-batch RSD of content uniformity. All runs were performed in triplicate, which required a total of 48 simulations. The simulation run order for each replicate of the design matrix was randomized. The age range for patient inclusion was not altered from that of the general population.

Table 2.5. Summary of the 2x2x2x2 full factorial experimental design.

Trial Number	Intra-Batch β	Inter-Batch β	Intra-Batch Content Uniformity	Inter-Batch Content Uniformity
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	1	1	0	0
5	0	0	1	0
6	1	0	1	0
7	0	1	1	0
8	1	1	1	0
9	0	0	0	1
10	1	0	0	1
11	0	1	0	1
12	1	1	0	1
13	0	0	1	1
14	1	0	1	1
15	0	1	1	1
16	1	1	1	1

2.2.10 Statistical Analyses

Sensitivity analyses were conducted on the results of the two DOEs using standard least-squares regression and an effects screening approach to determine the factors that had a significant influence on the risk to inefficacy and toxicity. This approach calculated the type III sums of squares. The inputs (Tables 2.4 and 2.5) were coded as nominal and the responses were coded as continuous. A full factorial model was initially generated to consider all potential interactions. Thereafter, fractional factorial models were assessed. Standard least-squares regression was also used to determine the final models for inefficacy and toxicity; both the inputs and responses were coded continuous. The significance level (α, not to be confused with the Weibull scale parameter) for all analyses was 0.05. All statistical analyses were conducted in Matlab (version 7.1, The MathWorks, Natick, MA) or JMP (version 8.0.1, SAS Institute Inc., Cary, NC). The risk scores for inefficacy and toxicity were analyzed independently.

2.3 Results and Discussion

2.3.1 Relationship between Toxicity and Inefficacy

It is important to understand the underlying relationship between inefficacy and toxicity in view of the fact that one of the fundamental objectives of the drug development process is to minimize the incidence of both adverse events. This can be accomplished using the PD model(s). Figure 2.9 illustrates the probabilistic relationship between inefficacy and toxicity, both of which are functions of theophylline concentration; Figure 2.9 is merely a 3-dimensional representation of the PD data that appear in Figure 2.7. As might be expected, the likelihoods for inefficacy and toxicity

are invariant at concentrations well below the MEC; the probability of experiencing an inefficacious event is high at low concentrations, whereas the probability associated with toxic events is minimal. As theophylline levels approach that which has been reported to have clinical effects (~ 5 mg/L), however, the probability of inefficacy sharply declines and does not begin to stabilize until roughly the mid-point of the therapeutic range. Conversely, the probability for toxic events is relatively constant until the middle of the therapeutic window, at which point the likelihood dramatically increases. The probability of inefficacy is nearly at its lowest value beyond the middle of the therapeutic range, indicating that patients continue to experience clinical outcomes (e.g., bronchodilation) while enduring the adverse event(s). The PD model also reveals that the probability of inefficacy will never be below approximately 22 %, which suggests that the drug will not offer clinical benefits for certain patients.

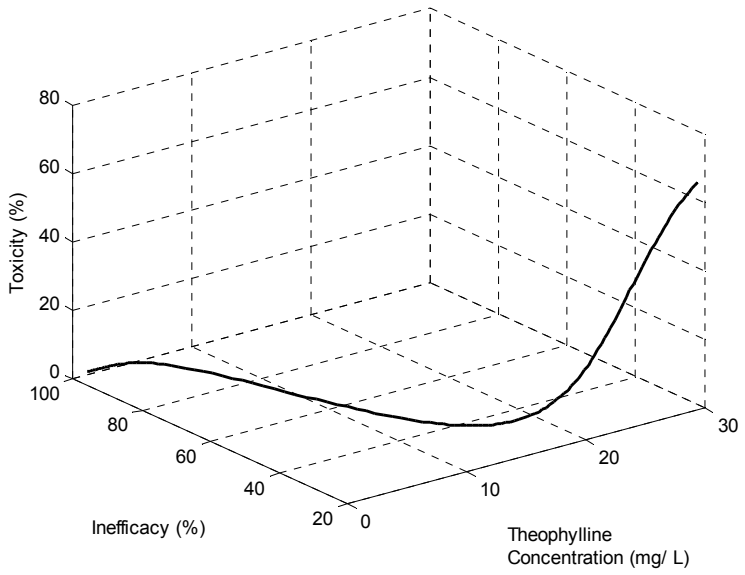

Figure 2.9. Plot of toxicity probability versus inefficacy probability versus theophylline concentration.

Figure 2.9 also illustrates that there are concentration ranges for which the two adverse events are inversely related to one anther. In particular, the probabilities for inefficacy and toxicity are linearly related with a correlation coefficient of –0.976 over 20 – 25 mg/L. Comparable negative correlations can be found locally over other concentration ranges. This is important since, depending on where patients are dosed, the interpretation of risk, for example, in a sensitivity analysis, has the potential to be analyzed in terms of either outcome without consideration of the other. Indeed, this should be verified before one of the two metrics is disregarded. While it is generally desirable to dose patients midway between the MEC and MTC, drug levels may consistently reside more close to one or the other. In this instance, a comparable change in both metrics may have more clinical relevance for one adverse event than it does for the other; this is more likely when the drug levels are near concentrations for which the relationship between inefficacy and toxicity deviates from linearity.

2.3.2 Dissolution Time Constant Optimization

The theophylline tablets produced from the three manufacturing routes resulted in various dissolution profiles. Thus, it was necessary to select an appropriate dissolution time constant that characterized the release of theophylline for the model dosage form prior to determining the conditional risk of product quality on clinical performance. Dissolution time constants ranging from 1 to 7 hours were assessed at 0.5 hour intervals since the majority of the tablets modeled yielded dissolution time constants in this range (Figure 2.4). Variability in the six manufacturing and patient compliance factors was prohibited during these trials. Each time constant was assessed in triplicate and the risk scores for inefficacy and toxicity are shown in Figure 2.10a and Figure 2.10b,

respectively. Ultimately, a dissolution time constant that minimized the risk to inefficacy and toxicity was desirable. Due to the local inverse relationship between inefficacy and toxicity, the minimum risk for inefficacy occurred at a time constant where risk of toxicity was the greatest. Therefore, a time constant of 5.0 hours was selected to characterize the model theophylline dosage form as this value favorably reduced the likelihood of toxic events observed at shorter time constants and concurrently minimized the increase in inefficacious events observed at longer time constants. These risk scores effectively represented the baseline risk from which variations in clinical performance were assessed. The remaining simulations were run using the parameter values indicated in Table 2.3 according to the experimental designs illustrated in Tables 2.4 and 2.5.

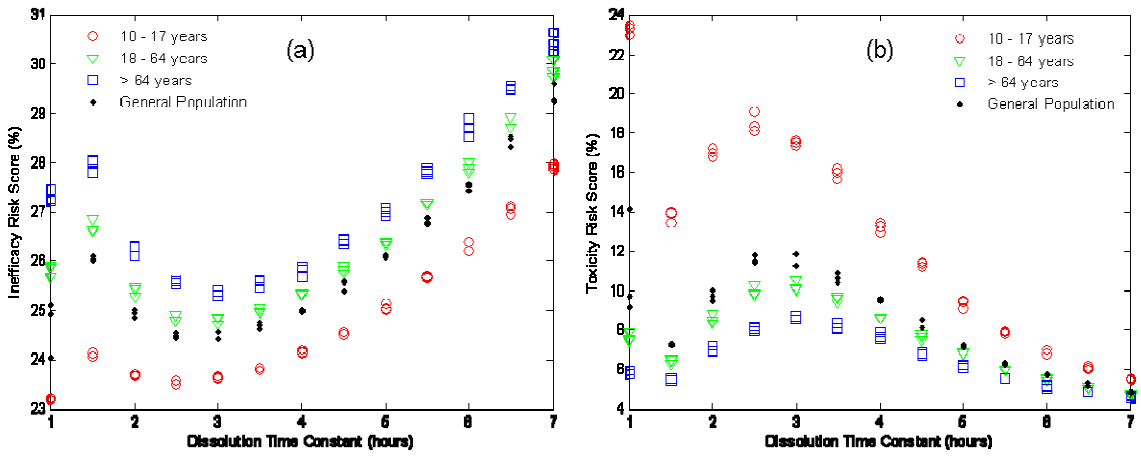

Figure 2.10. Plots of inefficacy (a) and toxicity (b) risk scores versus various dissolution time constants tested in different age-restricted sample populations.

2.3.3 2x2x2x2x2x2 Full Factorial Experimental Design

The following screening and modeling efforts utilized standard-least squares regression. The general approach to linear modeling assumes that the response is continuous over the range of negative infinity to positive infinity. This assumption can

be particularly problematic for proportional responses (e.g., probabilities), since, due to model error, the predictions can be outside of the anticipated range (e.g., 0 to 1 or 0 to 100). Therefore, the estimates and standard errors were examined for consistency and accuracy to substantiate the use of linear regression. The validity of the other assumptions of linear regression (e.g., homoscedasticity, linearity, normality) was also verified. Additionally, predictor variables were analyzed for multicollinearity.

The resultant inefficacy and toxicity risk scores for the 2x2x2x2x2x2 full factorial experimental design are summarized in Figure 2.11; quantile and other statistical metrics are also presented in Table 2.6. Risk scores for both inefficacy and toxicity were approximately unimodally distributed; the assumption of normality, therefore, is not unreasonable. Accordingly, transformations were deemed to be unnecessary. Simple linear regression revealed that the scores for inefficacy and toxicity were negatively correlated ($r = -0.997$). The inverse relationship was a direct result of the PD model and the dosing regimen; 95 % of the patients were dosed such that the CDF was consistently interpolated at theophylline concentrations of 20 – 25 mg/L (recall the inverse relationship of the PD model in this concentration range, which is illustrated in Figure 2.9). Given their inverse relationship, the discussion is predominately focused on toxicity. The corresponding inverse statistical relationships for inefficacy were confirmed.

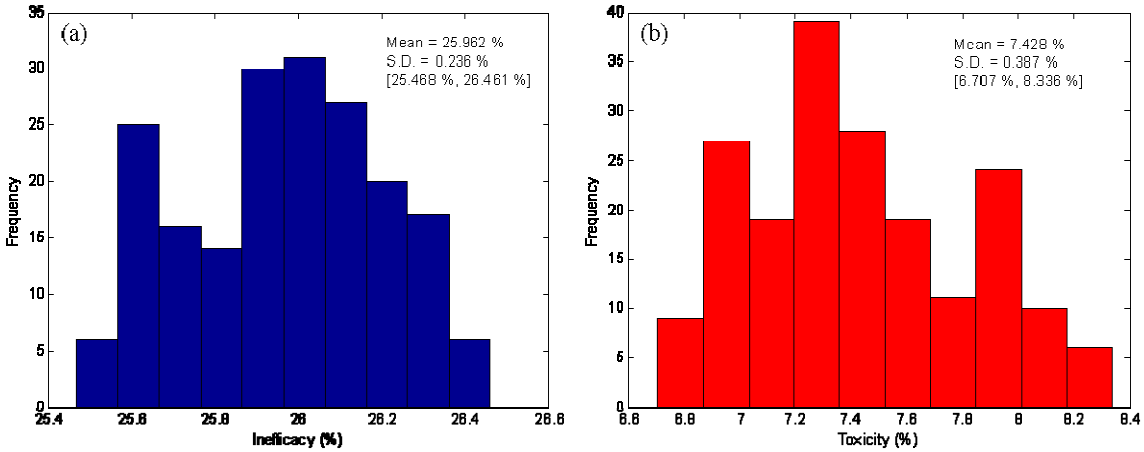

Figure 2.11. Frequency histograms of the resultant inefficacy (a) and toxicity (b) risk scores for the 2x2x2x2x2x2 full factorial experimental design (n = 192). The mean, standard deviation (S.D.), and range [,] of each adverse event are also provided.

Table 2.6. Summary statistics for the 2x2x2x2x2x2 full factorial experimental design.

Inefficacy				
Percentile	**Metric**	**Probability (%)**	**Metric**	
100.0%	maximum	26.461	Mean (%)	25.962
99.5%		26.461	Std Dev (%)	0.236
97.5%		26.369	Std Err Mean (%)	0.017
90.0%		26.281	Upper 95% Mean (%)	25.996
75.0%	quartile	26.135	Lower 95% Mean (%)	25.928
50.0%	median	25.979	Number of Observations	192
25.0%	quartile	25.785		
10.0%		25.629		
2.5%		25.538		
0.5%		25.468		
0.0%	minimum	25.468		
Toxicity				
Percentile	**Metric**	**Probability (%)**	**Metric**	
100.0%	maximum	8.336	Mean (%)	7.428
99.5%		8.336	Std Dev (%)	0.387
97.5%		8.184	Std Err Mean (%)	0.028
90.0%		7.998	Upper 95% Mean (%)	7.484
75.0%	quartile	7.703	Lower 95% Mean (%)	7.373
50.0%	median	7.373	Number of Observations	192
25.0%	quartile	7.137		
10.0%		6.934		
2.5%		6.820		
0.5%		6.707		
0.0%	minimum	6.707		

Sensitivity analyses were conducted to determine which factors had a significant impact on the risk of an adverse event. Figure 2.11 illustrates that the ranges of risk scores were narrow for both adverse events. Nevertheless, the full factorial screening revealed that three main effects, intra-batch RSD of content uniformity, rate of patient compliance, and standard deviation of the dosing interval, significantly influenced probability of experiencing a toxic event. In addition to these main effects, the first-order interaction between the rate of patient compliance and the standard deviation of the dosing interval was identified as significant. Two other higher order interactions were significant; however, they were determined to be spurious based on the insignificance of the other main effects that comprised the interaction terms. It is important to note that the same three main effects were determined to be significant for inefficacy. The interaction between the rate of patient compliance and the standard deviation of the dosing interval, however, was not strong enough to significantly alter the likelihood of an inefficacious event. This demonstrates the sensitivity of the risk simulation platform to asymmetric risk, a phenomenon that would go undetected with a standard "quality" metric such as Cpk, which does not account for clinical outcomes. Two additional higher order interactions were also significant for inefficacy; they were determined to be spurious as well.

Following the full factorial screening exercise, a 2^{nd} degree fractional screening was carried out to re-assess the main effects and first-order interactions. Analogous to the previous screening study, three main effects, intra-batch RSD of content uniformity, rate of patient compliance, and standard deviation of the dosing interval, as well as the first-order interaction between the rate of patient compliance and the standard deviation

of the dosing interval significantly influenced the probability of toxicity. Likewise, intra-batch RSD of content uniformity, rate of patient compliance, and standard deviation of the dosing interval significantly influenced the probability of inefficacy.

Subsequently, standard least-squares regression was used to compare several potential linear models. Ultimately, the final model for inefficacy included three main effects, intra-batch RSD of content uniformity, rate of patient compliance, and standard deviation of the dosing interval, while the model for toxicity included these three main effects and the first-order interaction between the rate of patient compliance and the standard deviation of the dosing interval. Intra-batch RSD of content uniformity and standard deviation of the dosing interval functioned to an increase in the probability of toxicity, whereas patient compliance decreased the likelihood of experiencing a toxic event.

Studentized residuals were, where appropriate, analyzed to verify that the assumptions of linear regression were valid for these data. One such plot, studentized residuals versus sample number, is provided in Figure 2.12. Additionally, plots of studentized residuals versus the predicted response values were examined, which did not suggest that these data were heteroscedastic (plots not shown). Abnormal patterns were not observed in any of the residual plots, which further substantiates the use of linear regression. The studentized residuals were also used to indentify outliers. The largest (absolute value of the) studentized residual for the toxicity model was 2.97, and a total of 11 residuals were above 2.0. Likewise, the largest (absolute value of the) studentized residual for the inefficacy model was 2.81 and, in all, 11 residuals were above 2.0. Therefore, no observations were removed for either model. The experimental design was

intentionally replicated to assess lack of fit. Testing of both models revealed that the null hypothesis, which stated that the model fit these data, could not be rejected. The final models for inefficacy and toxicity are summarized in Tables 2.7 and 2.8, respectively. The predicted versus measured plots for the two clinical outcomes, which also illustrate the appropriateness of the straight-line model, are shown in Figure 2.13.

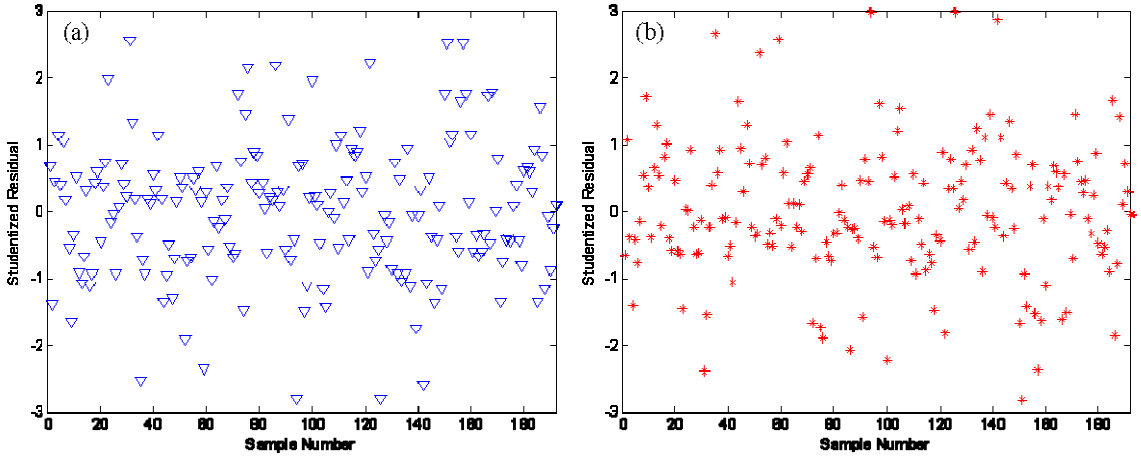

Figure 2.12. Plots of studentized residuals versus sample number for inefficacy (a) and toxicity (b) for the finalized linear models of the 2x2x2x2x2x2 experimental design.

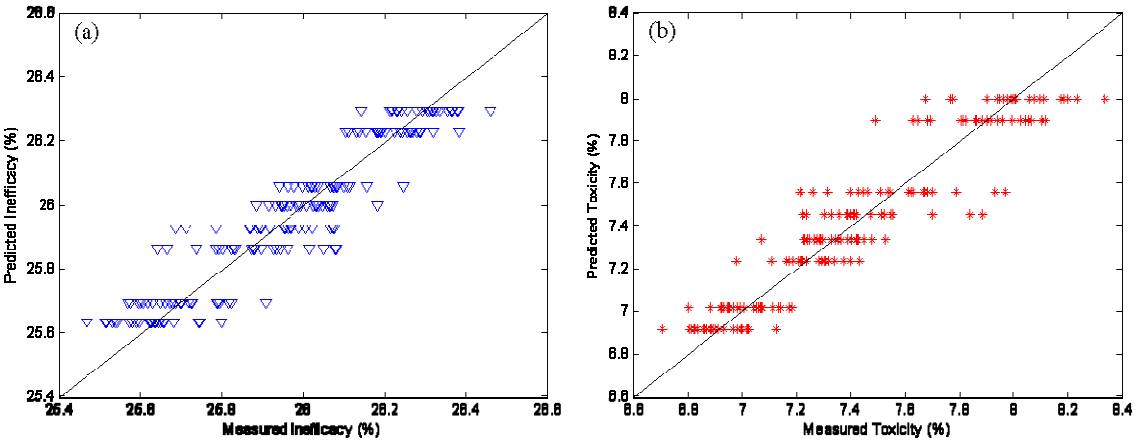

Figure 2.13. Predicted versus measured plots for the inefficacy (a) and toxicity (b) linear models. The unit line is shown in black.

Table 2.7. Final model results for inefficacy for the 2x2x2x2x2x2 experimental design.

Inefficacy				
Summary of Fit				
R^2	0.86851			
R^2 Adj	0.86641			
Root Mean Square Error	0.086			
Mean of Response	25.962			
Observations	192			
Analysis of Variance				
Source	**D.F.**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	3	9.225	3.075	413.917
Error	188	1.397	0.007	**Prob > F**
Total	191	10.621		1.54E-82
Lack of Fit				
Source	**D.F.**	**Sum of Squares**	**Mean Square**	**F Ratio**
Lack of Fit	4	0.020	0.005	0.664
Pure Error	184	1.377	0.007	**Prob > F**
Total Error	188	1.397		6.17E-01
				Max Rsq
				0.87038
Parameter Estimates				
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob > |t|**
Intercept	26.059	0.012	2094.647	0.0
Intra-Batch CU Variability	-0.061	0.012	-4.931	1.79E-06
Patient Compliance	0.234	0.012	18.792	4.97E-45
S.D. Dosing Time	-0.366	0.012	-29.399	3.13E-72

Table 2.8. Final model results for toxicity for the 2x2x2x2x2x2 experimental design.

Toxicity				
Summary of Fit				
R^2	0.86239			
R^2 Adj	0.85944			
Root Mean Square Error	0.145			
Mean of Response	7.428			
Observations	192			
Analysis of Variance				
Source	**D.F.**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	4	24.726	6.182	292.969
Error	187	3.946	0.021	**Prob > F**
Total	191	28.672		2.38E-79
Lack of Fit				
Source	**D.F.**	**Sum of Squares**	**Mean Square**	**F Ratio**
Lack of Fit	3	0.059	0.020	0.926
Pure Error	184	3.887	0.021	**Prob > F**
Total Error	187	3.946		4.29E-01
				Max Rsq
				0.86443
Parameter Estimates				
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob > |t|**
Intercept	7.270	0.021	346.746	1.37E-264
Intra-Batch CU Variability	0.100	0.021	4.783	3.49E-06
Patient Compliance	-0.381	0.021	-18.160	3.85E-43
S.D. Dosing Time	0.597	0.021	28.489	6.06E-70
Patient Compliance x S.D. Dosing Time	-0.116	0.042	-2.755	6.44E-03

To further scrutinize the final models, the 95 % confidence intervals for the expected mean value were grouped by all possible combinations of the independent variables (Table 2.9). Examination of the mean 95 % confidence intervals revealed that no two intervals overlapped across all possible input combinations. This was the case for both inefficacy and toxicity. Lack of overlap further underscored the significant change in risk scores induced by intra-batch RSD of content uniformity, rate of patient compliance, and standard deviation of the dosing interval. The data in Table 2.9 were

also used to generate plots of the predicted probabilities for inefficacy and toxicity adjusted for intra-batch content uniformity variability, patient compliance, and the standard deviation of the dosing interval (Figure 2.14). The interaction between the rate of patient compliance and the standard deviation of the dosing interval for the toxicity model is clearly demonstrated by the non-parallel nature of the lines in subplots c and d; subplots a and b substantiate the lack of interaction for the inefficacy model.

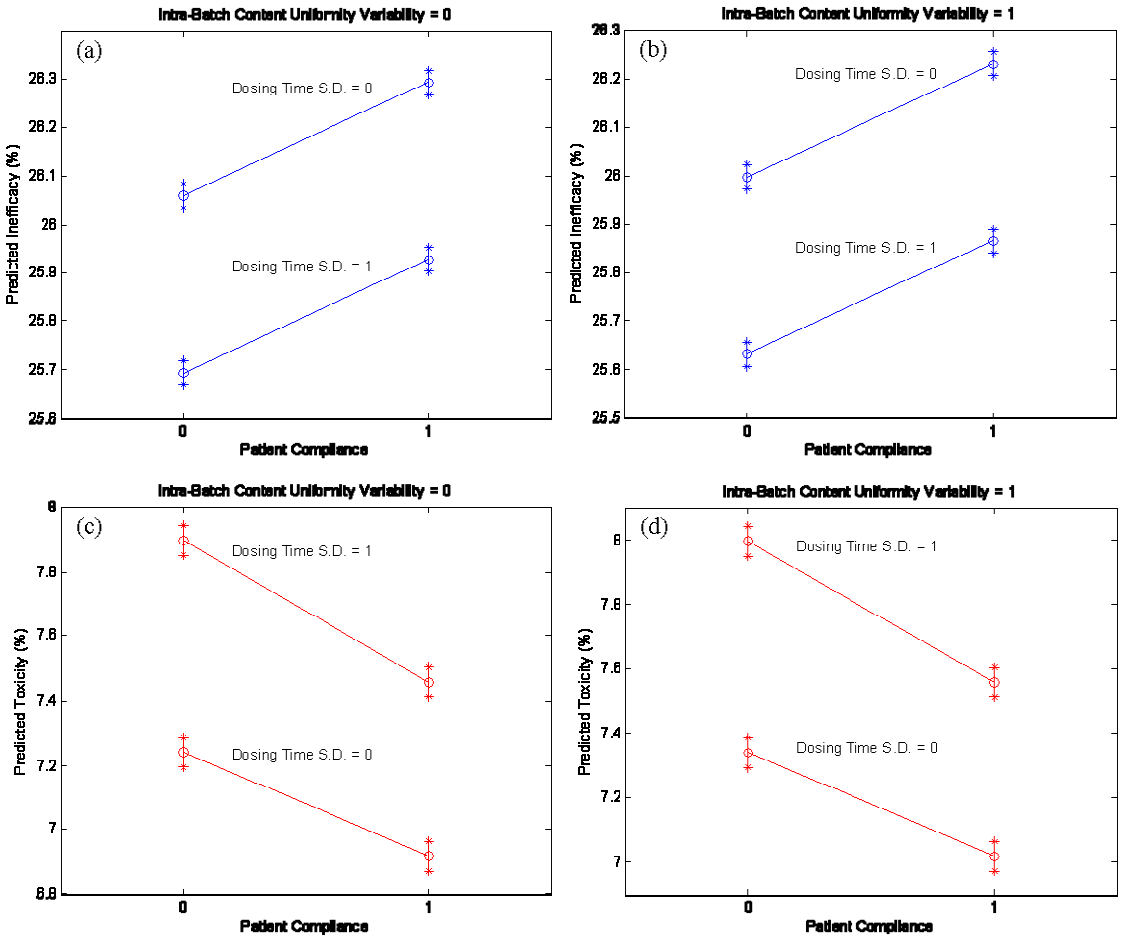

Figure 2.14. Plots of the predicted mean probabilities for inefficacy (a,b) and toxicity (c,d) adjusted for the effects of intra-batch content uniformity variability, patient compliance, and dosing time standard deviation. Asterisks denote the upper and lower values of the mean confidence intervals whereas the open circles represent the mid-point of the intervals.

Table 2.9. 95 % confidence intervals for the expected mean value grouped by all possible combinations of the independent variables. The acronym CI stands for confidence interval.

Inefficacy					
Intra-Batch Content Uniformity Variability	Patient Compliance	Dosing Time Standard Deviation		Predicted Inefficacy Mean 95 % CI	
				Lower	Upper
0	0	0		26.034	26.083
0	1	0		26.268	26.317
0	0	1		25.668	25.717
0	1	1		25.902	25.951
1	0	0		25.973	26.022
1	1	0		26.207	26.256
1	0	1		25.607	25.656
1	1	1		25.841	25.890
Toxicity					
Intra-Batch Content Uniformity Variability	Patient Compliance	Dosing Time Standard Deviation	Patient Compliance x Dosing Time Standard Deviation	Predicted Toxicity Mean 95 % CI	
				Lower	Upper
0	0	0	0	7.195	7.287
0	1	0	0	6.872	6.964
0	0	1	0	7.850	7.942
0	1	1	1	7.411	7.504
1	0	0	0	7.295	7.387
1	1	0	0	6.972	7.064
1	0	1	0	7.950	8.043
1	1	1	1	7.512	7.604

2.3.4 2x2x2x2 Full Factorial Experimental Design

Due to the overpowering variance explained by patient compliance and dosing time variability, a second experimental design was executed to evaluate the effects of manufacturing variability when patient compliance was 100 % and all doses were administered precisely at the scheduled dosing times. This was done to ensure that the two patient factors (at the levels assessed) did not mask subtle, yet important,

manufacturing effects. The resultant inefficacy and toxicity risk scores for the 2x2x2x2 full factorial experimental design are summarized in Figure 2.15; quantile and other statistical metrics are also presented in Table 2.10. Data were analyzed in a manner comparable to that in the 2x2x2x2x2x2 experimental design. As was observed in the 2x2x2x2x2x2 experimental design, the scores for inefficacy and toxicity were negatively correlated (r = -0.999). Likewise, the inverse relationship was a direct result of the PD model and the dosing regimen; 95 % of the patients were dosed such that the CDF was consistently interpolated at theophylline concentrations between 20 and 25 mg/L (recall the negative correlation between inefficacy and toxicity within this range). Consequently, the discussion is predominately focused on toxicity. The corresponding inverse statistical relationships for inefficacy were confirmed.

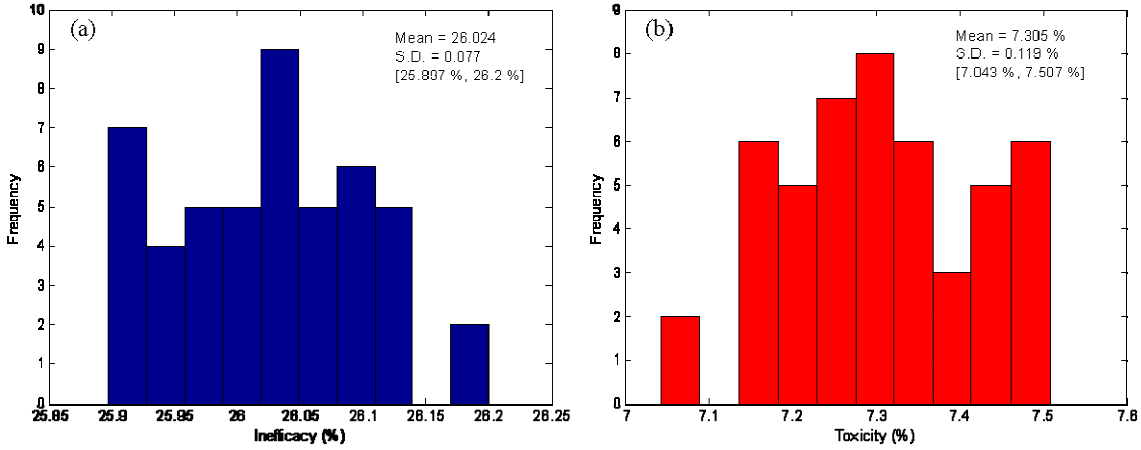

Figure 2.15. Frequency histograms of the resultant inefficacy (a) and toxicity (b) risk scores for the 2x2x2x2 full factorial experimental design (n = 48). The mean, standard deviation (S.D.), and range [,] of each adverse event are also provided.

Table 2.10. Summary statistics for the 2x2x2x2 full factorial experimental design.

Inefficacy					
Percentile	**Metric**	**Probability (%)**		**Metric**	
100.0%	maximum	26.200		Mean (%)	26.024
99.5%		26.200		Std Dev (%)	0.077
97.5%		26.196		Std Err Mean (%)	0.011
90.0%		26.125		Upper 95% Mean (%)	26.046
75.0%	quartile	26.086		Lower 95% Mean (%)	26.002
50.0%	median	26.025		Number of Observations	48
25.0%	quartile	25.962			
10.0%		25.911			
2.5%		25.897			
0.5%		25.897			
0.0%	minimum	25.897			
Toxicity					
Percentile	**Metric**	**Probability (%)**		**Metric**	
100.0%	maximum	7.507		Mean (%)	7.305
99.5%		7.507		Std Dev (%)	0.119
97.5%		7.507		Std Err Mean (%)	0.017
90.0%		7.483		Upper 95% Mean (%)	7.340
75.0%	quartile	7.401		Lower 95% Mean (%)	7.270
50.0%	median	7.301		Number of Observations	48
25.0%	quartile	7.208			
10.0%		7.151			
2.5%		7.048			
0.5%		7.043			
0.0%	minimum	7.043			

The results of the 4-factor experimental design were similar to those of the 6-factor design; intra-batch content uniformity was the only manufacturing factor that significantly affected risk of toxicity, even when deviations from the dosing regimen were not permitted. As was observed previously, intra-batch content uniformity was positively correlated with the change in toxicity risk scores and negatively correlated with inefficacy. Linear regression, however, did not yield models of considerable predictive power ($R^2 \approx 0.26$), which was most likely a consequence of the large standard deviation relative to the narrow range of the resultant risk scores (Table 2.10). Nonetheless, intra-

batch content uniformity is important in explaining the variation in inefficacy and toxicity risk scores.

It is important to note that the conditional risk, regardless of whether or not patients are compliant, is dependent upon the manufacturing estimates tested. For example, assume that the RSD of content uniformity (both inter- and intra-batch) is comparable to the estimates assessed (Table 2.3), but, due to poor control during tableting, the estimates for dissolution time variability are worse. Lack of control during tableting could result in highly variable compression pressures, which, in turn, would sequentially yield erratic (1) radial tensile strengths, (2) dissolution profiles, and (3) exposure-response profiles. These changes would undoubtedly affect the portion of variability explained by the inter- and intra-batch dissolution time constant factors.

The conditional risk is also expected to vary from product to product. While dissolution variability (at the level tested) did not significantly impact clinical performance for the model solid oral dosage system, it may very well significantly influence, for example, an immediate release tablet. For instance, moderate dissolution variability could result in sub-therapeutic levels at the critical time period following administration (e.g., 30 minutes), which would most likely result in clinical inefficacy. These effects were not as pronounced in the model system, most likely because the factors were assessed once patients were at steady-state. Dissolution variability, therefore, was not large enough to induce an adverse event.

In addition to product-dependence, risk to clinical performance is also dependent on the production method. A substantial change in the manufacturing route, such as from direct compression to wet granulation, is likely to considerably alter drug dissolution

(Figure 2.4), and, therefore, clinical performance if the change is not optimized with respect to the desired QTPP. The adjustment, however, does not need to be so dramatic to have an effect on the patient. A switch in the blending protocol from a v-blender to a bin blender is likely to affect the inter- and intra-batch content uniformity if the PCCPs are not optimized with regard to the QTPP. Likewise, a formulation modification from intra- to extra-granular addition of the granulating binder has the potential to alter drug dissolution, and ultimately, inefficacy or toxicity. For these very reasons, changes in the manufacturing protocol should be investigated with regard to their impact on clinical performance. This can be accomplished by directly linking the process to clinical performance via a design space.

2.3.5 Risk Simulation: A Piece of the Modernization Puzzle

The risk simulations were conducted at the point in time where actual metrics of the manufacturing characteristics were available (Table 2.3). Nonetheless, this was not designed to imply that a risk assessment can only be initiated once the manufacturing variability metrics are accessible. Given an approach such as the one described herein, a risk assessor has the opportunity to assume values for parameters and/or attributes he/she believes have the potential to influence risk. Since these are approximations, uncertainty can then be propagated through the platform to gain a better understanding of which factors significantly affect risk. As the actual values become available, they can be incorporated, and the risk assessment can then be repeated.

Similarly, finalized/optimized components of the risk simulation platform (Figure 2.1) are not necessary to conduct the assessment. This work was completed using a hypothetical PD model. That does not mean, however, that the additional assessments

cannot be carried out once a legitimate probabilistic PD model of theophylline is generated. Moreover, gender- and age-specific estimates for the usage of barbiturates and benzodiazepines were assumed. As actual data become available, the patient population can be re-generated, at which point the risk simulations can be repeated. Components, whether they are as substantial as the PK or PD model, or as small as a single coefficient in a model, can be replaced as better estimates, or the finalized elements, are arrived at. The risk simulation platform was constructed to be modular to provide this flexibility. This iterative procedure can be implemented throughout the drug development process to enhance product and process understanding.

One of the objectives of the Critical Path Initiative is to accelerate the time-to-market of innovative, safe, and effective medical products by changing the approach to product development. Sponsors are encouraged to utilize innovative techniques to investigate the manufacturability, safety, and efficacy of candidate molecules and/or drug products.[9] This objective can certainly be expanded to include approaches that examine the impact that changes, such as those instituted through comparability protocols, have on the manufacturability, safety, and efficacy of currently marketed products. The multivariate risk simulation platform used in this work provides the opportunity to simultaneously study the effects of manufacturing, compliance, and physiologic and pathophysiologic states on the safety and efficacy of drug delivery systems. This is true for new chemical entities and previously marketed drugs alike.

The multivariate risk simulation platform also serves as a resource allocation tool, which can help fulfill the public health objective of offering affordable medications. For example, analysis of the model drug system revealed that intra-batch content uniformity

was the only manufacturing factor assessed that significantly influenced the probability of an adverse event. Risk simulation, therefore, identified intra-batch content uniformity as a CQA. While the other factors are not to be disregarded (beyond the range evaluated), it would be unreasonable to a invest large sum of resources into further reducing the precision of manufacturing such that dissolution variability consistently passed strict specifications seeing as how the current level of variability did not significantly alter clinical performance. Such an investment would needlessly inflate the overall product cost. The manufacturer should still be cognizant of the insignificant factors, however, since additional levels of variability in one or more of the inconsequential variables could elicit a significant change in clinical performance. This could be accomplished by monitoring and controlling their variability within the limits of acceptable risk to clinical performance (i.e., quality, as it is redefined). Resources should, on the other hand, be devoted to understanding and controlling the PCCPs for intra-batch content uniformity such that risk of adverse events is minimized.

Up to this point, the utility of the risk simulation platform has largely been couched on harnessing explicit patient and product knowledge to evaluate clinical performance (i.e., quality) as it relates to pharmaceutical production. As was discussed in the literature survey (refer to section 1.3.3.2), simulation has played an important role in clinical trials. A risk simulation approach such as this one also has the potential to contribute greatly in this area. Despite the fact that conditional risk was investigated using the general population, the risk simulation platform can also delineate sub-populations that display disparate risk levels (Figure 2.10). This supports the selection of participants for inclusion in clinical trials, with the ultimate objective of reducing the

likelihood of the drug being toxic or ineffective. Although drugs that are capable of being safely and effectively administered to the general population are desirable, certain patient factors often preclude individuals from taking a given medication. These sub-populations must be quickly identified so as to allow safe treatment. The gamut of patient factors that interact to affect drug action will not always be available initially; however, data from drugs of the same class or defensible estimates can be used as starting points. Subsequent clinical trial data can then be integrated within the simulation platform to better understand the conditions that predispose patients to adverse clinical outcomes. Once validated, these data can then be used to carefully market the product.

Whether launching a clinical trial or beginning treatment in a doctor's office, the risk simulation platform can facilitate arriving at a safe and effective individualized dose based on the volunteer/patient's ascertainable factors (e.g., age, gender, BMI, smoking and drinking status, known concomitant drugs). Together with the acting physician's expertise, the likelihood of adverse events can be minimized before the individual is ever administered the drug. With the appropriate data, this methodology would eliminate the oftentimes cyclic dose (by weight or some other dosing nomogram), monitor (serum/plasma levels), and adjust (as needed) approach which unquestionably jeopardizes the health of the individual if the first attempt is inaccurate. Admittedly, any uncertainty present in the risk simulation platform would also endanger the individuals. The platform, however, could be validated through a randomized clinical trial (simulation supervised versus unsupervised dosing) thereby mitigating the effects of unmodeled variance.

Thus far, the inter-relationship of the risk simulation components was utilized none other than to generate risk scores. These links are illustrated by the solid arrows in Figure 2.1. The risk scores, however, can be harnessed to oversee and/or optimize certain components (dotted arrows). For example, the dosing guidelines (whether for the general population or select sub-populations) can be adjusted to minimize the risk of adverse events. Furthermore, feedforward and feedback manufacturing controls can be instituted (via process and control models) to control PCCPs such that the desired level of clinical performance is attainted. Similarly, raw material variability can be integrated such that the process can be adjusted to compensate for risk imparted by incoming components. Chapter 3 uses the risk simulation platform to generate a design space for the model solid oral dosage system that is bounded by risk scores. Once the design space has been created, control models can be developed to ensure that production is maintained at a level of acceptable risk. Since risk scores are continuous, one or more acceptance thresholds must be arrived at. This should be a multidisciplinary decision that weighs factors such as feasibility, cost to the consumer, and risk-to-benefit ratios.

2.4 Conclusions

A risk simulation platform that integrated population statistics, drug delivery system characteristics, dosing guidelines, patient compliance estimates, production metrics, and PK, PD, and IVIVC models to investigate the impact of manufacturing variability on clinical performance of a model theophylline solid oral dosage system was described. This work was predicated on requests to re-define pharmaceutical quality in terms of risk by linking production characteristics to clinical attributes. Manufacturing precision was characterized by inter- and intra-batch content uniformity and dissolution

variability metrics, while clinical performance was described by a probabilistic PD model that expressed the probability of inefficacy and toxicity as a function of theophylline plasma concentrations. At the levels assessed, both patient compliance variables, percent of doses taken and dosing time variability, significantly impacted risk of inefficacy and toxicity. In addition to these factors, intra-batch content uniformity variability elicited a significant change in risk scores for the two adverse events, and, therefore, was identified as a CQA. This is the first in a series of chapters that demonstrate how pharmaceutical quality can be recast to explicitly communicate risk as it relates to clinical performance. Future research will focus on constructing a design space that directly links critical process parameters to quantitative estimates of inefficacy and toxicity risk. Thereafter, control models can be developed to supervise production such that clinical performance of the final product is within the hyperspace.

Chapter 3: Performance-Based Quality Specifications: The Relationship between Process Critical Control Parameters, Critical Quality Attributes, and Clinical Performance

3.1 Introduction

The quality of pharmaceutical products is currently evaluated through a series of independent tests (e.g., USP <711> and <905>) that do not explicitly communicate the clinical consequences of product variability. Univariate specifications disregard potential multivariate and nonlinear interactions that affect risk of clinical performance.[20] For example, a clinical inter-dependence between API content and drug release is expected to exist for solid oral dosage systems. Super-potent tablets with elevated release rates compromise patient safety due to increased drug levels in the blood. Furthermore, such product poses a specific (toxic) risk to patients whose drug clearance rates are suppressed (e.g., alcoholics with severe liver damage). Therefore, under the current testing paradigm, it is conceivable that in-specification (i.e., passing) product could, in certain patients, pose a greater clinical risk than product determined to be out-of-specification (Figure 3.1). If quality were to be redefined by linking production characteristics to clinical attributes, however, specifications for product release could be then be established on the basis of clinical risk.

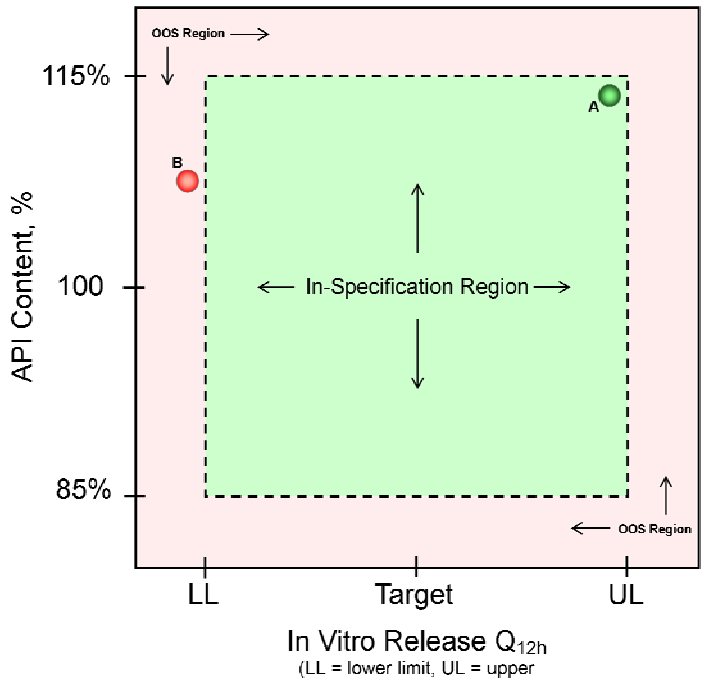

Figure 3.1. Schematic illustrating final product release testing with USP <905> and <711>. Given the specifications, only tablet A would pass the release tests, despite the fact it may very well pose a greater risk than tablet B. OOS stands for out-of-specification. Figure reproduced from the work of Cogdill and Drennen.

Besides failing to directly communicate clinical consequences, these tests dichotomize quality as satisfactory or unacceptable, despite the fact that the unit-to-unit quality can be disparate in a passing lot. While equally weighting the risk of all samples in and of itself diminishes their utility, the validity of these very tests, which are used to gauge the quality of pharmaceutical products, is now being questioned. A recently published article evaluated the sensitivity of the USP <905> test for content uniformity.[119] To conduct the investigation, lots containing varying degrees of non-conforming material were simulated *in silico*. Lots of different distributions (i.e., normal, log normal, bimodal, and uniform) were also tested. The authors concluded that USP <905> "is relatively insensitive to detecting non-conforming material." Furthermore,

despite that fact that the simulated lots contained considerable portions of out-of-specification material, the test did not consistently fail these lots until the percentage of defects was in excess of 20 %. This underscores the potential ineffectiveness of the USP <905> and the corresponding final product specifications in gauging quality. Similar arguments, such as the one framed by Tsong *et al*.,[120] have also questioned the sensitivity of USP <711> for detecting lots where considerable portions of the tablets are out-of-specification.

Under the current (empirical) paradigm for product development, specifications are often established on the basis of achievable levels of process reproducibility long after the production method is selected from feasibility and profitability criteria. For instance, a particular production facility might be equipped for, and decidedly experienced in (and, therefore, biased towards) wet granulation methods. The drug-specific process would have been inevitably transferred from R&D to production, at which point a considerable level of process understanding is available. The final product specifications would more than likely be set in parallel with (or following) the scale-up efforts. These specifications seek to maximize production yield while concurrently minimizing foreseeable undesirable outcomes (e.g., product recalls, adverse patient reactions). Since the majority of these outcomes have profound direct (rework costs, legal fees) and indirect (brand image) monetary connotations, companies often prefer to impose excessively strict specifications to mitigate the effects of process and/or product uncertainty (i.e., their effect on clinical performance). Thus, there exists a need to directly link process and product variability to clinical performance to maximize safety, effectiveness, and affordability.

In contrast to the empirical paradigm, a first-principles approach to product development would determine the acceptable limits of clinical performance before consideration of the production method. Process development would commence only after the performance-based quality specifications were defined. In other words, production would be tailored to achieve the clinical specifications without prior knowledge of process reproducibility. The idea of a first-principles approach might be further extended to begin with a particular disease state and a preliminary understanding of the patients with said condition. Next, a pharmacophore would be optimized for a specific receptor, which collectively elicit pharmacologic action. Using available patient, PK, and PD knowledge, acceptable levels of clinical performance would then be proposed, but should be flexible to integrate incoming clinical data (i.e., uncertainty in initial knowledge). The tolerable limits for clinical performance will then be used to condition the manufacturing process and evaluate product quality. Ideally, a drug delivery system and its associated manufacturing process would be selected on its ability to achieve the highest levels of clinical performance; however, companies may very well be restricted by prior investments (e.g., purchase of specific production equipment). The process, irrespective of the delivery system ultimately implemented, can be optimized with respect to clinical performance. Once the delivery system has been identified, the CQAs, or the physical, chemical, biological, or microbiological properties or characteristics that affect clinical performance (i.e., quality), can then be identified. Thereafter, the process variables that ultimately impact clinical performance (i.e., PCCPs) can be isolated. The first-principles approach to product development is summarized in Figure 3.2.

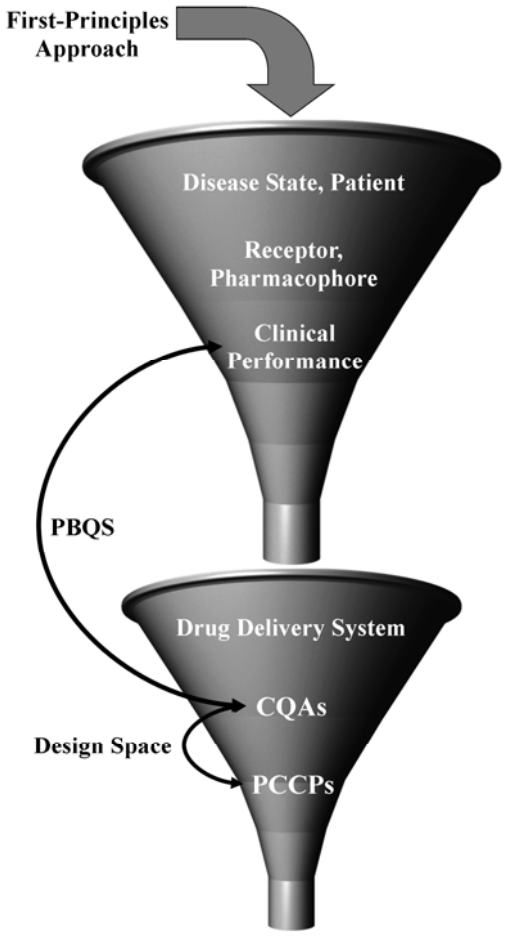

Figure 3.2. Schematic diagram illustrating the first-principles approach for product development.

Design space has assumed a predominant role in the pharmaceutical industry. It is, therefore, appropriate to discuss the impact that recasting quality in terms of clinical performance has on components associated with the idea of design space. In its current form, design space is considered "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality."[12] As the appendices in ICH Q8R1 illustrate, the hyperspace is defined by PCCPs, which thereby condition the CQAs. Consequently, the design space is specific to a unit operation or a single production process. It defines the operational range of process parameters known to influence

(surrogates of) product quality (e.g., dissolution). Design space can be thought of as the link between PCCPs and CQAs (Figure 3.2). MacGregor and Bruwer recently suggested such an architecture where design space would characterized the combination of CQAs that offered product whose quality was compliant with specific safety and efficacy requirements; however, they did not propose a means to directly connect CQAs and clinical performance (i.e., quality).[121]

If quality was to be considered in terms of risk by linking production characteristics to clinical attributes, design space would be, as MacGregor and Bruwer have alluded to, an n-dimensional interaction of inputs that have been demonstrated to ensure clinical performance. CQAs would remain fundamental to the design space; however, they would become inputs to, rather than outputs of, the hyperspace. The definition of a CQA would effectively remain unchanged, but the physical, chemical, biological, or microbiological properties or characteristics would be directly related to clinical performance (e.g., inefficacy or toxicity risk scores) rather than indirectly linked via surrogate indicators of quality (e.g., moisture content, dissolution, friability). The identification and utilization of a CQA, therefore, would reflect the re-standardization of "quality." Likewise, the definition of a PCCP would not need to be amended. Conceptually, however, a PCCP would be intimately related to clinical performance rather than the surrogate marker. Performance-based quality specifications (PBQS), therefore, are proposed to be the link between CQAs and clinical performance (Figure 3.2). As such, they can be used to define a design space centered on quantitative estimates of clinical performance.

Following the first-principles approach (Figure 3.2), construction of the design space would be hierarchical. At the outset, the underlying relationship between product attributes (e.g., API content, drug release rate) and clinical performance would define the space. If the drug delivery system has already been identified (e.g., uncoated tablet), a risk simulation platform (reference Chapter 2) could conceivably be used to generate the initial hyperspace before a single unit is ever produced. Nevertheless, once the production method has been identified, process and product knowledge, in conjunction with experimental design, risk assessment, simulation, and, if desired, PAT, could be harnessed to identify the attributes that are critical to clinical performance (i.e., CQAs). In contrast to the current design space methodology, a hyperspace that defines the relationship between CQAs and clinical performance is extremely powerful since it is not specific to a given process. Theoretically, one of several processes (e.g., direct compression, wet granulation, roller compaction) could be integrated within the design space since, within reason, the CQAs remain invariant from process to process. While partial transferability between, for example, production lines for an extended release solid oral dosage form is feasible, universal transferability (e.g., direct integration of process knowledge for a transdermal drug delivery system) would most likely be unachievable since different CQAs would be expected to influence clinical performance. Transfer of the process-independent design space to a specific production line would be accomplished by identifying the process parameters that affect clinical performance, and, subsequently generating process models that effectively serve as transfer functions between the PCCPs and CQAs. The final design space, which would then be subject to regulatory discretion, would illustrate the underlying relationship between PCCPs and

clinical performance. Control models could also be integrated to adjust the process (for incoming raw material variability, in-process product variability, environmental fluctuations, etc.) such that clinical performance of the final product lies within the design space.

It is apparent that re-defining pharmaceutical quality in terms of risk by relating process and patient characteristics to clinical performance is beneficial for the manufacturer, the regulator, and the patient alike. While the revamp is long overdue, standardizing such a fundamental concept alters current pharmaceutical quality initiatives. This work addressed the design space modifications that accompany the new definition of pharmaceutical quality. The objectives were to (1) illustrate the potential drawbacks of the current paradigm for solid oral dosage form release testing, (2) generate artificial design spaces for a model theophylline solid oral dosage system that are conditioned on quantitative estimations of inefficacy and toxicity risk, and (3) address the potential of a release paradigm where the quality of pharmaceutical products can be evaluated with regards to clinical performance.

3.2 Materials and Methods

3.2.1 Risk Simulations

The risk simulator, which has been described previously (refer to Chapter 2), was used to conduct *in silico* studies. Several modifications were made to the simulator to alter its functionality. However, unless noted otherwise, the theophylline regimen was not adjusted. In the previous studies, the mean Weibull dissolution time constant was optimized to minimize global risk. The underlying Weibull dissolution time constant

distribution was then generated using this mean and estimates of the inter- or intra-batch RSD. The resultant distributions were randomly sampled to yield the test parameter for the current simulation iteration; a new time constant was sampled for each dose administered. Since the purpose was not to evaluate constrained manufacturing scenarios, but rather to perturb the simulation inputs beyond the levels tested in the initial risk assessments, particularly for dissolution variability given that it did not have a significant impact on clinical performance (reference Chapter 2), the dissolution time constants were fixed for each risk simulation trial.

Additionally, rather than randomly sampling the variability in theophylline uniformity from its underlying distributions, deviation from label claim was specified for each risk simulation trial. For a given risk simulation trial, individual patient doses were estimated using an iterative dosing scheme (refer to section 2.2.7). All doses administered for a given trial, however, were adjusted to deviate by the same percentage from the nominal amount. Analogous to the previous studies, each simulation trial consecutively tested patients until the variability of the risk estimates for both inefficacy and toxicity were below the difference threshold of 10^{-4}. Furthermore, the absolute change was required to retain a value below the threshold for 250 successive patients before the simulator converged on the risk estimates. All doses were assumed to be taken at their scheduled times; treatment was administered every 12 hours for 30 days. The age range for the general population was not restricted. A total of 288 independent simulation trials were run.

Data pertaining to 12 distinct wet granulation batches of the model theophylline tablets were utilized to conduct these simulations (refer to section 2.2.3). As was

described earlier, 12 and 10 units per batch were sampled for dissolution and content uniformity testing, respectively. Prior to dissolution or content uniformity testing, a reflectance spectrum for both sides of each tablet was acquired over the wavelength range of 1100 – 2498 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 5000-II, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD). Each dissolution profile was modeled using a two-parameter Weibull function, which yielded a shape and scale parameter. Data corresponding to tablets that underwent content uniformity testing were used to construct calibration models for the prediction of theophylline.

Calibration models were constructed using NIR reflectance data, expressed in absorbance units, of the 120 tablets (240 spectra) and amount of theophylline (%) as determined via UV-Vis spectroscopy (refer to section 2.2.3). Partial least-squares regression[122] was used via the SIMPLS algorithm[123] to relate spectroscopic response to theophylline amount. Preprocessing routines, including standard normal variate (SNV) scaling, detrending, derivatives, and combinations of the preceding were tested.[124] The most favorable data pretreatment method was selected based on a minimization of cross-validation error. Contiguous block cross-validation with a block size of 5 was used to generate the temporary cross-validation models. For all preprocessing routine(s) employed, spectroscopic data were mean-centered while reference data were scaled to zero mean and unit variance. Model rank was chosen as the point where a rapid decline in the incremental variance captured was observed, cognizant of the expected feasible limit of dimensionality based on the factors varying within the design. Detrending (first-order) was determined to be the optimal data pretreatment method. Two latent variables were necessary to adequately model these NIR absorbance data for the prediction of

theophylline. This calibration model was subsequently applied to the 288 spectra of the 144 tablets subjected to dissolution testing in order to predict their theophylline levels (i.e., testing data set); predicted amounts were compared to the nominal levels for each tablet as per the original design.[117] Predicted and reference values were used to determine the root-mean-standard error (RMSE). The RMSE for cross-validation (RMSECV), calibration (RMSEC), and testing (RMSET) were calculated using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3.1}$$

where y_i is the measured amount, $\hat{y}_i$ is the predicted amount, and n is the number of samples for the data set under consideration. Summary statistics for the calibration model are provided in Table 3.1.

Table 3.1. Calibration summary statistics for the prediction of theophylline.

Data Type	NIR (Reflectance)
Method	PLS
Preprocessing	First-Order Detrending
Latent Variables	2
Component	Theophylline
R^2_{CAL}	0.980
R^2_{CV}	0.948
R^2_{TEST}	0.982
RMSEC (%)	0.96
RMSECV (%)	1.57
RMSET (%)	1.04
$Bias_{CV}$ (%)	-0.248
$Bias_{TEST}$ (%)	-0.250

Projection of the spectra of the tablets subjected to dissolution testing onto the calibration model provided theophylline amounts. These data supplement the available

information content (i.e., Weibull shape and scale parameters) for these tablets. Predicted theophylline percentages for the same tablet varied slightly due to minor inconsistencies (e.g., positioning, heterogeneity, constituent packing) detected between the two tablet surfaces. While the theophylline percentages were unique for a given tablet, the Weibull scale and shape parameters were identical for duplicate scans. Therefore, 288 distinct responses were available for the 144 tablets.

With the percent theophylline and Weibull scale and shape parameter data now available, each distinct response was used as an input set for a given risk simulation trial. Specifically, predicted theophylline percentages were compared to the nominal values to calculate the percent deviation from label claim; this percentage was then used to adjust each patient's dose once it was individualized using the iterative dosing scheme. The Weibull parameters and deviation from label claim were constant throughout a given simulation; the inputs were updated with values corresponding to a new tablet (or the reverse side of the same tablet) only once the current trial met the convergence criteria. Thus, the only element varying within a given trial was the pool of patients tested; all patients received tablets that displayed identical content uniformity and dissolution characteristics.

3.2.2 USP Testing

Several USP monographs for theophylline products have been published in the USP-NF.[125] Tests for uniformity and dissolution draw on the general chapters <905> and <711>, respectively, where the individual monographs indicate the specifications for each test for a particular product. A monograph for theophylline tablets is available; however, it is limited to immediate-release products. Since the model drug delivery

system is an extended-release tablet, specifications were extracted from its closest analog, the monograph for extended-release capsules; where appropriate, the specifications for products labeled for dosing every 12 hours were used. This assumption is justified given that the specifications outlined in the monograph ought to have been optimized in terms of biopharmaceutical performance. Since theophylline is highly soluble and highly permeable, comparable dissolution curves should yield similar safety and efficacy profiles for the two delivery systems. As stated in the monograph, extended-release capsules are to contain between 90 and 110 percent of the labeled amount of anhydrous theophylline. All content uniformity analyses were conducted using this criterion. Theophylline content was determined using the aforementioned calibration model. Given that dissolution of the model tablets was conducted in deionized water (refer to section 2.2.3), dissolution profiles were evaluated using Test 10; none of the other 9 tests are applicable.[126] Acceptance criteria for the percentage of label claim dissolved at the specified times are summarized in Table 3.2. All dissolution testing was conducted using these criteria according to USP <711>.[80]

Table 3.2. Acceptance criteria for the percent label claim of theophylline dissolved at the specified times according to USP Test 10 for 12-hour extended-release capsules.

Time (hours)	Amount Dissolved (% Label Claim)	Designation
1	6 - 27	Q_1
2	25 - 50	Q_2
4	65 - 85	Q_4
8	> 80	Q_8

Dissolution profiles (% theophylline released versus time) of the 144 tablets manufactured via the wet granulation method were evaluated according to the criteria outlined in Table 3.2 using a program written in-house. The program determined the

percent theophylline dissolved (i.e., Q_1, Q_2, Q_4, and Q_8) at the specified times (i.e., 1, 2, 4, and 8 hours) for each tablet. Once these data were available, each tablet was evaluated according to USP <711> under the assumption that the one tablet was representative of a uniform lot (i.e., all tablets within the given lot had identical Q_1, Q_2, Q_4, and Q_8 values). Level once acceptance criteria for extended-release dosage forms states that all Q values for each unit tested must fall within the stated ranges. Furthermore, all units tested must have a Q value greater than the amount specified for the final test time.[80] Since the lots are assumed to be uniform, proceeding to level two testing upon failure of level one or level three upon failure of level two will not alter the test outcome. Thus, the tablets were only subjected to level one testing according to the criteria outlined in Table 3.2.

Given that the risk simulator was constructed to process Weibull parameters estimates rather than Q values, least-squares regression was used to relate the dissolution time constants ($T_{63.2}$) of the 144 tablets manufactured via the wet granulation method to Q_1, Q_2, Q_4, and Q_8; linear as well as polynomial fits were evaluated. The predictive models for Q_1, Q_2, Q_4, and Q_8 are summarized in Table 3.3; RMSEC was determined using Equation 3.1. Since $T_{63.2}$ is a reasonable surrogate for Q_1, Q_2, Q_4, and Q_8, at least inasmuch as the scenarios tested, USP <711> results for the model system were displayed in terms of $T_{63.2}$. Reported specifications for $T_{63.2}$ only encompass tablets (lots) that passed USP <711> according to the criteria outlined in Table 3.2.

Table 3.3. Summary of the predictive models for Q_1, Q_2, Q_4, and Q_8 with $T_{63.2}$ as the predictor.

Dissolution Time (hours)	Designation	Transformation	Equation	RMSEC (%)	R^2_{CAL}
1	Q_1	$1/Q_1$	$1/Q_1 = 0.00017*T_{63.2}^3 - 0.0031*T_{63.2}^2 + 0.026*T_{63.2} – 0.0030$	2.58	0.874
2	Q_2	None	$Q_2 = -0.34*T_{63.2}^3 + 6.65*T_{63.2}^2 - 45.58*T_{63.2} + 129.75$	1.91	0.987
4	Q_4	$1/Q_4$	$1/Q_4 = 1.68x10^{-5}*T_{63.2}^4 - 0.00037*T_{63.2}^3 + 0.0028*T_{63.2}^2 – 0.0054*T_{63.2} + 0.013$	1.80	0.993
8	Q_8	None	$Q_8 = 0.28*T_{63.2}^3 - 4.60*T_{63.2}^2 + 15.88*T_{63.2} + 84.53$	1.29	0.987

3.3 Results and Discussion

3.3.1 USP <905> and <711> 'Testing'

The scores for inefficacy and toxicity for the 288 trials were negatively correlated ($r = -0.8785$), which is not as strong as that which was observed in the risk assessments (reference Chapter 2). The reduction in covariance between inefficacy and toxicity is a consequence of assuming uniform production; lots which deviate to a greater extent from nominal result in theophylline plasma concentrations that are consistently off-target, and, therefore, outside the concentration range where inefficacy and toxicity are inversely related. Due to the lack of (complete) correspondence, results were depicted for both adverse events.

While it would have been straightforward to assign artificial values for theophylline content and the Weibull dissolution parameters, it was particularly desirable to ascertain the true relationship between these factors seeing as how they are confounding. Thus, to avoid testing a combination of inputs that were potentially impractical, production tablets from 12 different batches were used to estimate theophylline content and the Weibull parameters. The 144 tablets manufactured via a wet granulation method were delineated into 288 distinct responses where each response included percent theophylline (hence, deviation from nominal or label claim), the Weibull shape and scale parameters, and Q_1, Q_2, Q_4, and Q_8. These data afford the opportunity to evaluate the tablets, which were assumed to each represent a uniform production lot, according to the USP tests for content uniformity and dissolution. Likewise, these data can be harnessed to investigate the impact that specific

manufacturing scenarios have on clinical performance. In other words, the risk simulator can be used to project the clinical consequences of producing lots displaying various content uniformities and dissolution characteristics.

The results of the 288 simulation trials were arranged according to the input values for theophylline, expressed as percent deviation from nominal, and the Weibull dissolution time constant. Figure 3.3 displays plots of the resultant inefficacy and toxicity risk scores versus API deviation from nominal (synonymous to deviation from label claim). The lower and upper specifications limits for content uniformity, as outlined in the individual monograph for extended-release theophylline capsules, are also depicted. All of the 288 lots tested were well within the acceptance limits for content uniformity. Similarly, Figure 3.4 displays plots of the resultant inefficacy and toxicity risk scores versus the Weibull dissolution time constant. The lower and upper specification limits for dissolution are also depicted. For the conditions assessed, tablets with dissolution time constants between 3.10 and 3.77 hours resulted in Q_1, Q_2, Q_4, and Q_8 values that met the acceptance criteria outlined in Table 3.2. In total, only 62 of the 288 lots passed USP <711>.

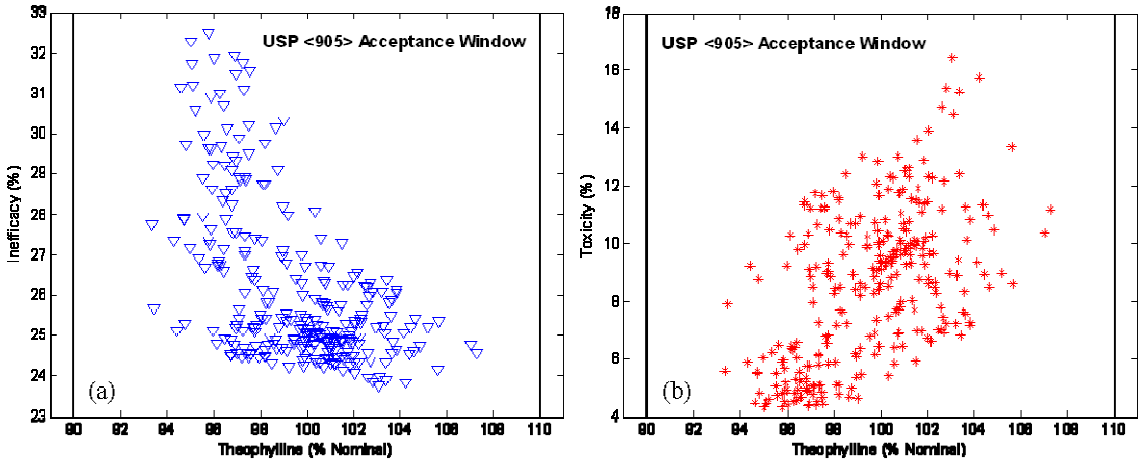

Figure 3.3. Plot of risk scores for inefficacy (a) and toxicity (b) versus theophylline amount, expressed as percent deviation from nominal (i.e., label claim). The bold vertical lines indicate the lower and upper specification limits for content uniformity analysis as per USP <905>.

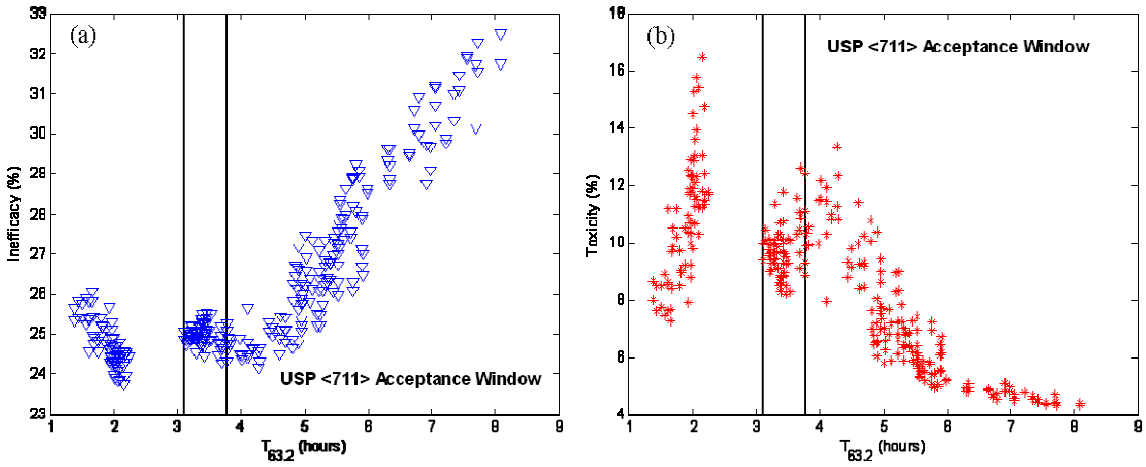

Figure 3.4. Plot of risk scores for inefficacy (a) and toxicity (b) versus the Weibull dissolution time constant. The bold vertical lines indicate the lower and upper specification limits for dissolution analysis as per USP <711> and the individual monograph for theophylline extended-release capsules.

USP tests for content uniformity and dissolution were conducted in sequential fashion, which is in line with the current approach for solid oral dosage form release testing. Combining the outputs of the USP tests and the risk simulations exposes one of the potential pitfalls of the current system. As Figure 3.3 illustrates, samples that met the criteria for content uniformity, criteria which are intended to ensure the safety and

efficacy of the final product, display disparate levels of clinical performance. The maximum risk scores observed in the risk assessments, for which content uniformity was identified as a CQA, were 26.46 % and 8.34 % for inefficacy and toxicity, respectively, well below many of the scores obtained during these studies. Despite considerable differences in clinical performance, tablets sampled from any of the simulated lots would have been taken as acceptable. It is quite presumptuous to uniformly categorize all 288 scenarios. Figure 3.4 offers a similar prospective on the current approach to final product testing. Although the lots which passed USP <711> had relatively low risk scores for inefficacy, the toxicity risk scores were exceptionally high, the majority of which were above the maximum risk score observed in the risk assessments (reference Chapter 2). Albeit an artificial construct, these data emphasize the potential disconnect between final product specifications and clinical performance. This underscores the importance of defining quality in terms of risk by linking production characteristics to clinical attributes.

Another potential shortcoming of the current approach to final product testing lies in its univariate approach. As previously noted, a clinical inter-dependence between API content and drug dissolution is likely. These data support such an interaction. As Figure 3.3 illustrates, lots that were low in theophylline content posed a higher risk for inefficacy, whereas lots that were in excess of label claim presented a greater likelihood of toxicity. Similarly, Figure 3.4 depicts the clinical impact of various dissolution time constants; smaller dissolution time constants (corresponds to faster release rates) result in lower probability of being inefficacious, while larger dissolution time constants (corresponds to slower release rates) pose a reduced risk of toxicity. The impact on the

risk scores is exacerbated when the effects of content uniformity and dissolution are considered simultaneously. Take, for instance, the group of toxicity risk scores that correspond to dissolution time constants between 1 and 2 hours (Figure 3.4b). Although a low dissolution time constant generally yields a high probability of toxicity, time constants between 1 and 2 hours resulted in a considerable range in risk scores. This can be explained, at least in part, by the interaction between content uniformity and $T_{63.2}$; super-potent tablets with short dissolution time constants posed a greater toxic risk than did sub-potent tablets. Although this interaction is reflected in the multivariate risk score, it is not accounted for in the univariate specifications utilized in USP <905> and <711>. This begs the question, how good are the current specifications at ensuring quality?

It is not entirely appropriate to question the validity of final product specifications on the basis of risk score magnitudes without ascribing some level of clinical significance to the scores themselves. While conclusions regarding the clinical significance of a risk score should be drawn by a multidisciplinary team comprised of clinicians, scientists, process engineers, statisticians, etc., thresholds defining low, medium, and high risk were instituted to further underscore the potential drawbacks of assessing final product quality under the current paradigm. Inefficacy risk scores less than or equal to 25.979 %, greater than 25.979 % and less than or equal to 26.461 %, and greater than 26.461 % were classified as low, medium, and high risk, respectively. Likewise, toxicity risk scores less than or equal to 7.373 %, greater than 7.373 % and less than or equal to 8.336 %, and greater than 8.336 % were classified as low, medium, and high risk, respectively. The median risk scores from the 2x2x2x2x2x2 full factorial experimental design (refer to Chapter 2) were used to delineate low and medium risk, while the maximum observed

risk scores for inefficacy and toxicity were used to distinguish medium and high risk for the two adverse events. Figure 3.5 illustrates the combined effect of theophylline content and $T_{63.2}$ after the scores were categorized according to low, medium, and high risk. The acceptance criteria for USP <905> and <711> are also defined. Figure 3.5a illustrates that these criteria are likely much too strict (more so for <711>) when considering inefficacy, and are inaccurate for assurance of safety. The criteria would ideally take into consideration both adverse events, and, therefore, would need to be optimized to simultaneously mitigate the likelihood of inefficacy and toxicity. Depending on the risk-to-benefit ratios determined by the multidisciplinary team, the two adverse events are likely to be disproportionally weighted when optimizing the criteria. For instance, certain toxic events (e.g., headache) might be acceptable (to both the clinician and the patient), thereby reducing the sensitivity of the test criteria to specific fluctuations in toxicity risk scores as a result of the corresponding increase in efficacy.

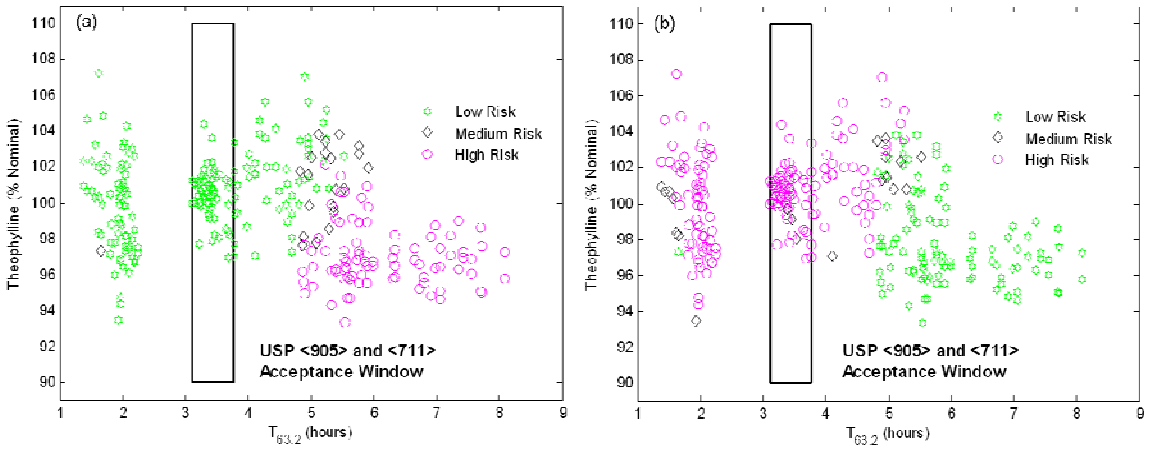

Figure 3.5. Plots of theophylline content versus Weibull dissolution time constant for inefficacy (a) and toxicity (b) with the clinical outcomes categorized according to low, medium, and high risk. The acceptance limits for USP <905> and <711> are indicated by the bold rectangle.

3.3.2 Design Space

In light of the deficiencies which limit assessments of final product quality, risk simulation has been used to delineate the relationship between the CQAs of the model drug delivery system and its clinical performance (i.e., PBQS). The relationship between theophylline content, $T_{63.2}$, and inefficacy and toxicity risk is depicted in Figure 3.6. As the following discussion addresses, the PBQS can then be used to generate a design space where the incoming CQAs have been conditioned on estimates of inefficacy and toxicity risk. To be consistent with the examples previously shown, clinical thresholds were instituted to generate the design spaces. Inefficacy risk scores less than or equal to 25.979 %, greater than 25.979 % and less than or equal to 26.461 %, and greater than 26.461 % were classified as low, medium, and high risk, respectively. Likewise, toxicity risk scores less than or equal to 7.373 %, greater than 7.373 % and less than or equal to 8.336 %, and greater than 8.336 % were classified as low, medium, and high risk, respectively.

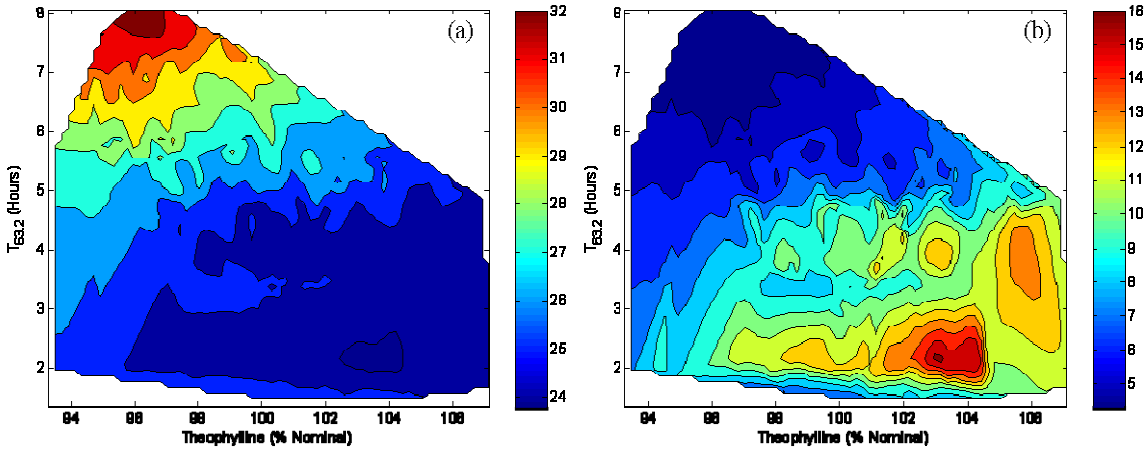

Figure 3.6. Contour plots depicting the relationship between theophylline content and the Weibull dissolution time constant for inefficacy (a) and toxicity (b). The color bars represent the ranges of the risk scores (%).

Two artificial design spaces (independently) conditioned on quantitative estimates of inefficacy and toxicity risk are presented in Figure 3.7. The example design spaces were confined to combinations which resulted in low risk scores. Although the design spaces were restricted to three dimensions (projected onto a two-dimensional contour plot) to facilitate viewing, the reader should make note that the hyperspace can be n-dimensional. The design spaces can be adjusted to reflect a multidisciplinary team's conclusion regarding acceptable risk using the data represented in Figure 3.6. Validation should confirm that design space appropriately discriminates clinical risk.

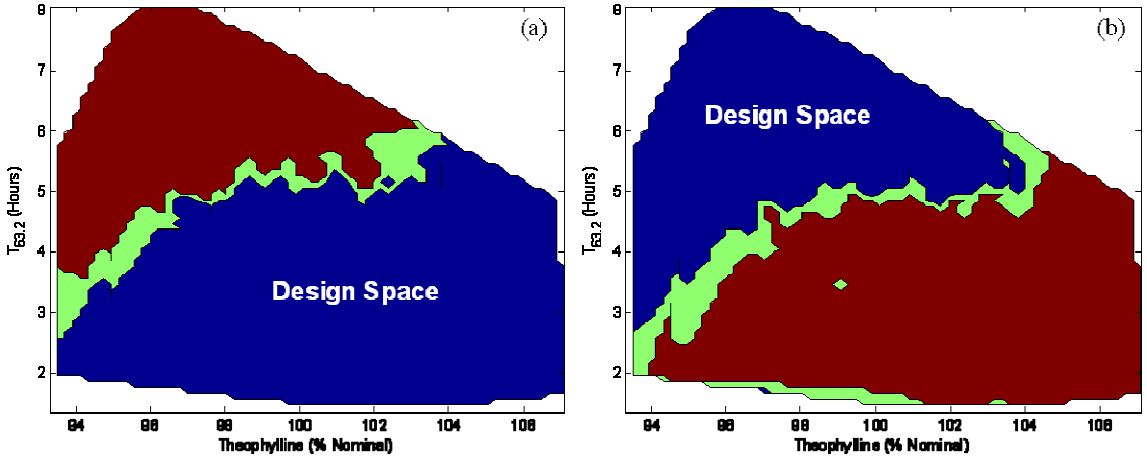

Figure 3.7. Simulation based design spaces for a model theophylline extended-release tablet conditioned on inefficacy (a) and toxicity (b) risk scores. Dark blue, sea green, and burgundy signify low, medium, and high risk, respectively.

Although dissolution variability was not identified as a CQA during the risk assessment, it was included as an input to the design space. At first glance, it might appear as though the results of the risk assessment were discounted in view of the fact that $T_{63.2}$ was an input to the design space. The results of the risk assessment, however, must not be taken at face value. While dissolution variability was not critical to quality (at $\alpha = 0.05$) within the experimental design, it interacts with content uniformity, which

was critical to clinical performance. If $T_{63.2}$ were not monitored, it may unknowingly drift to a level that would, when combined with the effects of in-specification content uniformity, unfavorably impact clinical performance. Thus, it is imperative that $T_{63.2}$ be included in the design space. This is true for all non-critical parameters that interact (within ranges reasonable for a given process) with CQAs.

Furthermore, there is no direct penalty (e.g., cost, model complexity) for including $T_{63.2}$ in the design space given that the architecture of the risk simulator (Figure 2.1) predetermined the mathematical relationship between the $T_{63.2}$ and inefficacy and toxicity risk. Integration of an attribute not previously studied in the risk assessment would, however, invalidate the PBQS. This underpins the need to combine the information gained from the risk assessment with knowledge of the process and product.

3.3.3 A New Release Paradigm

A design space conditioned on probabilistic estimates of clinical performance has significant regulatory implications; the advantages currently promoted (e.g., reduced comparability protocol reporting) are still applicable. Although hyperspaces that utilize CQAs as inputs are informative, those which directly link PCCPs to clinical performance (i.e., those which are specific to a process) are likely to offer the greatest regulatory flexibility; this is largely a function of their direct interpretability regarding the process itself. Once the PBQS are established, efforts to generate process models can commence; data upon which process models are developed may, in fact, already be available. These models could serve as transfer functions between the PCCPs and CQAs, effectively linking a given process to clinical performance. Control models, therefore, can be instituted to oversee process parameters such that clinical performance is maintained

within the design space. Although material attributes (e.g., API particle size distribution) were not included in the artificial construct, they are by no means precluded from such an approach. Material attributes can be integrated into the risk assessment as well as the process and control models.

Thus, a design space constructed from PCCPs (as well as material attributes) and estimates of clinical performance can be, coincident with the current intentions of the initiative, subjected to regulatory review. The appropriate validation data, which should include actual clinical trial and production data, can be used to demonstrate the predictive power of the risk simulator, the process and control models, and, ultimately, the process-specific design space. Regulatory approval would, therefore, create a potential platform for real-time product release. In essence, final product conforming to the design space need not be re-evaluated for assurance of quality. The design space, which has its own integrated specifications, thereby, serves as the ultimate final product release test. In spite of the best design of experiments, there always exists the potential for introduction of unmodeled variance (e.g., change in raw material supplier, shift in production sites). Therefore, the design space (and all of its associated models) must be continually managed, and updated as necessary.

3.4 Conclusions

In silico simulations were conducted to generate inefficacy and toxicity risk scores for 288 uniform lots of extended-release theophylline tablets displaying explicit content uniformity and dissolution variability. These data were used to demonstrate potential weaknesses of the univariate specifications utilized in USP <905> and <711> tests for content uniformity and dissolution, respectively. The simulated results

underscore several potential deficiencies. First, in-specification product demonstrated a large range of quality. This was especially true for the content uniformity results, where all of the 288 uniform lots assessed conformed to the final product specifications. Secondly, dissolution testing, which was conducted using criteria outlined in USP <711> and the individual monograph for extended-release theophylline capsules, revealed that certain out-of-specification lots posed a lower risk of inefficacy and toxicity. Lastly, the relationship between final product specifications and clinical performance was obfuscated even further when the specifications for USP <905> and <711> were considered simultaneously in the context of acceptable/unacceptable levels of inefficacy and toxicity risk. The simulated results illustrated that the criteria for dissolution testing were too strict for inefficacy and were inaccurate for toxicity. This was related to the tests' inabilities to account for interaction between content uniformity and dissolution variability.

This work also addressed, principally for design space, the consequences of re-defining pharmaceutical quality in terms of risk by linking production characteristics to clinical attributes. A risk simulator was used to define the underlying relationship between quality attributes and clinical performance for the model theophylline extended-release tablets. Both critical (i.e., content uniformity) and non-critical (i.e., Weibull dissolution time constant) attributes were used as inputs to the design space, which was conditioned on quantitative estimates of inefficacy and toxicity risk. Such a design space can then be applied to specific processes using process models, which relate PCCPs to the quality attributes, as transfer functions to ultimately link process parameters to clinical performance. The direct link enhances the information content of the design

space by omitting quality surrogates (e.g., dissolution, moisture content) that are utilized in current design space practices. Design spaces conditioned on estimates of clinical performance may ultimately expedite real-time product release efforts by moderating final product testing.

Chapter 4: A Near-Infrared Spectrscopic Investigation of Relative Density and Crushing Strength in a Four-Component Solid Dosage System

The material presented in Chapter 4 was previously published in Short SM, Cogdill RP, Wildfong PLD, Drennen III JK, Anderson CA 2009. A near-infrared spectroscopic investigation of relative density and crushing strength in four-component compacts. Journal of Pharmaceutical Sciences 98(3):1095 - 1109.

4.1 Introduction

Near-infrared spectroscopy (NIRS) has demonstrated its utility for the analysis of intact pharmaceutical dosage systems.[124] Given the quantitative capabilities when used in conjunction with multivariate calibration, NIRS is frequently employed for the non-destructive prediction of constituent concentrations within pharmaceutical compact and tablet matrices. It is well understood, however, that near-infrared (NIR) spectra convey information pertaining to both the chemical and physical nature of the samples.[127] Signals related to physical variation (e.g., hardness or crushing strength) are commonly treated as interferences in composition calibration models. Generally, variations in physical factors such as relative density (solid fraction) result in a characteristic baseline shift,[128-133] the effect of which can be suppressed by mathematical treatment. Two common chemometric algorithms used for this purpose are standard normal variate (SNV)[134] scaling and multiplicative scatter correction (MSC).[135] Suppression of these physical features usually reduces the number of model factors required to achieve optimum performance.

Beyond chemical quantification, numerous studies have documented the modeling of NIR data for the characterization of physical attributes. One aspect of solid oral dosage forms that has been examined using this technology is tablet hardness or crushing strength. Drennen and Lodder pioneered the use of NIRS for the measurement of tablet hardness.[129,136] Subsequently, numerous publications have demonstrated NIR calibrations for crushing strength.[128-133,137-149] The majority of these articles suggest that an increase in tablet hardness results in a smoother tablet surface, increasing apparent NIR absorption (presumably because more light is lost to specular reflectance). Otsuka and Yamane took a unique approach in which they generated calibration models to predict the eventual hardness of tablets produced at constant compaction pressure from powder mixtures having varying blend times (hence, changing the distribution of constituents).[138] While the authors were able to generate calibration models having significant correlation to tablet hardness, they were unable to relate spectral changes to a particular constituent. The authors suggested that the NIR calibration was detecting not only composition, but more subtle factors including porosity, pore structure, and the tablet surface and geometry.[138] Three other groups have investigated the use of NIRS for the analysis of tablet porosity; all determined that NIRS was suitable for the measurement of tablet porosity, reporting varying levels of success with the use of different mathematical techniques.[141,143,149]

The objectives of this work were to demonstrate (1) characteristic absorption effects of NIR radiation by compacts of varying relative density and crushing strength, (2) the source of spectral variability resulting from varying relative density and crushing strength, (3) how multivariate analysis can be used to elucidate the effects of chemical

composition upon the physical parameters of pharmaceutical solid oral dosage forms, and (4) how calibration experimental design influences spectroscopic variance. Finally, these data are used to present a revised rationalization for NIR sensitivity to compact hardness.

4.2 Materials and Methods

4.2.1 Compact Production

The details regarding the production of the compacts used for this work have been described elsewhere.[150] Briefly, a fully-balanced, four-constituent mixture design consisting of anhydrous theophylline (Lot No. 92577, Knoll AG, Ludwigshafen, Germany), Lactose 316 Fast Flo NF Monohydrate (Lot No. 8502113061, Hansen Labs, New Berlin, WI), microcrystalline cellulose (MCC, Avicel PH-200, Lot No. M427C, FMC BioPolymer, Mechanicsburgh, PA), and soluble starch GR (Lot No. 39362, EMD Chemicals Inc., Gibbstown, NJ) was generated. The approximate median particle size of the theophylline, lactose, MCC, and starch (reported by documentation from their respective suppliers), was 90, 100, 180, and 17 microns, respectively. No further analyses or operations were performed on the raw materials to determine or alter their particle size or distribution. Twenty-nine design points were chosen to cover a wide composition range and to remove any possibility of factor aliasing (Table 4.1). The mixture covariance matrix demonstrates that the design is balanced in all factors, giving equal emphasis to all constituents.

Table 4.1. Concentration design for the 4-component compacts.

Design Point	Anhydrous Theophylline (w/w)	Lactose Monohydrate (w/w)	MCC (PH200) (w/w)	Soluble Starch (w/w)
1	0.600	0.200	0.200	0.000
2	0.400	0.400	0.200	0.000
3	0.200	0.600	0.200	0.000
4	0.400	0.200	0.400	0.000
5	0.200	0.400	0.400	0.000
6	0.200	0.200	0.600	0.000
7	0.600	0.200	0.000	0.200
8	0.400	0.400	0.000	0.200
9	0.200	0.600	0.000	0.200
10	0.600	0.000	0.200	0.200
11	0.400	0.200	0.200	0.200
12	0.200	0.400	0.200	0.200
13	0.000	0.600	0.200	0.200
14	0.400	0.000	0.401	0.200
15	0.200	0.200	0.400	0.200
16	0.000	0.400	0.400	0.200
17	0.200	0.000	0.600	0.200
18	0.000	0.200	0.600	0.200
19	0.400	0.200	0.000	0.400
20	0.200	0.400	0.000	0.400
21	0.400	0.000	0.200	0.400
22	0.200	0.200	0.200	0.399
23	0.000	0.400	0.200	0.400
24	0.200	0.000	0.400	0.400
25	0.000	0.200	0.400	0.400
26	0.200	0.200	0.000	0.600
27	0.200	0.000	0.200	0.600
28	0.000	0.200	0.200	0.600
29	0.250	0.250	0.250	0.250

Materials for each design point mixture were dispensed by weight (Data Range, Model No. AX504DR, Mettler Toledo, Columbus, OH), and subsequently transferred to 25 mL glass scintillation vials. In total, 6000 mg of material was weighed out for each point, and the nominal weights for all constituents were adjusted to the observed mass data to calculate actual concentration. The vials were mixed for 5 minute cycles by placing them on the rotating drive assembly of a Jar Mill (US Stoneware, East Palestine, OH, USA). After each blending period, a NIR reflectance spectrum was acquired

through the bottom of each vial (FOSS NIRSystems 5000, FOSS NIRSystems, Inc., Laurel, MD), and an *ad hoc* partial least-squares II (PLS-2) calibration was constructed to assess homogeneity. Mixtures were assumed to be homogeneous when further mixing failed to yield an increase in the calibration coefficient of determination.

The mixtures from each design point were then subdivided and compacted at one of 5 pressures (67.0, 117.3, 167.6, 217.8, 268.1 MPa) using a Carver Automatic Tablet Press (Model No. 3887.1SD0A00, Wabash, IN) equipped with 13 mm flat-faced punches and die. A dwell time of 10 seconds was employed. Six compacts weighing approximately 800 mg were produced per design point, with the sixth tablet's compaction pressure pseudo-randomly selected from one of the five possible levels, for a total of 174 compacts. The compaction order was randomized to minimize heteroscedastic errors. Following compaction, the samples were retained in the sealed vials for 15 days prior to spectroscopic analysis.

Compacts consisting of each pure component were produced in a similar manner. Approximately 800 mg of each component was compacted at 9 different compaction pressures (67.0, 90.5, 117.3, 140.8, 167.6, 191.0, 217.8, 241.3, and 268.1 MPa) using the same press and tooling. Four additional pressures were employed to increase the number of data points in each compaction profile. Three replicate compacts were produced at every compaction pressure for each constituent, yielding 27 pure compacts per material. The manufacturer's lot of lactose monohydrate used for the compaction profiles differed from that used to make the 4-component compacts (Lot No. 8505010961, Hansen Labs, New Berlin, WI); all other materials were from the aforementioned lots.

4.2.2 Data Acquisition, Instrumentation, and Software

Near-infrared reflectance measurements were acquired for both sides of each compact (excludes pure component compacts) using a scanning monochromator instrument, equipped with a Rapid Content Sampler, over the wavelength range of 1100 – 2498 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 5000-II, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD). Two *ad hoc* partial least-squares II (PLS-2) calibrations, using the constituent concentrations as reference data, were constructed from reflectance spectra corresponding to a specific surface of the tablets. Since the coefficients of determination did not differ until the third decimal place, it was decided to only consider measurements for one compact face.

Transmittance measurements were acquired on a scanning monochromator instrument equipped with an InSight^{TM} Tablet Analyzer over the wavelength range of 600 – 1898 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 6500, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD). The wavelength range of the transmittance spectra was truncated to 800 – 1400 nm due to limitations imposed by the sample pathlength.

All spectral data were analyzed in the Matlab environment (version 7.1, The MathWorks, Natick, MA) using the PLS_Toolbox (version 3.0, Eigenvector Research, Inc., Manson, WA) and software developed at Duquesne University.

4.2.3 Physical Testing

Following compaction, the samples were stored in sealed glass scintillation vials and were removed only for analysis. After an approximate 40 day span to allow for radial expansion, the compacts were weighed (Data Range, Model No. AX504DR,

Mettler Toledo, Columbus, OH) and their thicknesses and diameters were measured using a digital micrometer (TESA Micromaster, Model No. IP54, Brown & Sharpe, North Kingstown, RI). Assuming a cylindrical geometry, these data were used to estimate each compact's density (ρ_{comp}) according to the formula

$$\rho_{comp} = \frac{m}{\pi \cdot \left(\frac{d}{2}\right)^2 \cdot t} \tag{4.1}$$

where m, d, and t are compact mass, diameter, and thickness, respectively. The pure component tablets were allowed to relax for approximately 55 days before their masses and dimensions were acquired. The true density of each constituent was estimated via helium pycnometry (Micromeritics Accupyc, Model No. 1330, Particle & Surface Sciences Pty. Limited, Gosford, New South Wales, Australia). The mean of five powder sub-samples was used for each constituent (Table 4.2). The true densities (ρ_{true}) of each compact were estimated using the equation

Table 4.2. Component true densities as determined by helium pycnometry.

Component	Theophylline	Lactose	MCC	Starch
True Density (g/cm^3)	1.41	1.5063	1.5084	1.477
True Density (g/cm^3)	1.4071	1.5072	1.5084	1.4768
True Density (g/cm^3)	1.4034	1.5072	1.508	1.4766
True Density (g/cm^3)	1.4024	1.5076	1.5053	1.4766
True Density (g/cm^3)	1.4012	1.5074	1.5058	1.4767
Average True Density (g/cm^3)	1.405	1.507	1.507	1.477
Standard Deviation (g/cm^3)	0.0036	0.0005	0.0015	0.0002

$$\rho_{true} = \sum_{i=1}^{n} X_i \cdot \rho_{true,i} \tag{4.2}$$

where X_i and $\rho_{true,i}$ are the w/w contribution and the true density for the i^{th} component within a n-component sample. For the work herein, n was either 4 or 1 corresponding to

the mixture and pure component compacts, respectively. The relative density (D) of each mixture and pure component compact was estimated as

$$D = \frac{\rho_{comp}}{\rho_{true}} \tag{4.3}$$

The crushing strength of the compacts, reported in kiloponds (kp), was estimated for the mixture compacts by a diametric crushing test (Vision Tablet Testing System, Model No. ElizaTest 3+, Elizabeth-Hata International, North Huntingdon, PA). The maximum recordable value for this particular instrument was 55.9 kp. Twelve calibration compacts and one test compact were evaluated to have values of (at least) 55.9 kp; one other test sample yielded a value of 0.0 kp. The information pertaining to these 14 compacts was withheld from the subsequent crushing strength modeling but was included in the relative density analyses.

4.2.4 Regression Analyses

Near-infrared reflectance and transmittance spectra were independently modeled in an identical manner. Partial least-squares (PLS) regression[122] was used via the SIMPLS algorithm[123] to relate spectroscopic response to relative density and crushing strength. No spectral preprocessing was employed; data were only transformed to absorbance (log (1/R) or log (1/T)) and subsequently mean-centered. All reference data (relative density and crushing strength) were scaled to zero mean and unit variance prior to modeling. Given that the calibrations were intended to model the physical variance within the spectra, preprocessing routine(s) were not applied. Certain applications may necessitate spectral preprocessing to suppress interfering signals (e.g., spectrometer drift); however, implementation of such methods may reduce the net analyte signal[150] of a

feature. For this work, preprocessing routines, including SNV scaling, detrending, derivatives, and combinations of the preceding, were tested at the outset;[124] all of these pretreatments reduced the ability to predict relative density and crushing strength from NIR reflectance and transmittance spectra.

The optimum model was selected based on minimization of "batch-wise" cross-validation error,[151] where the batches in this instance corresponded to the 29 different concentration levels. The root-mean-standard error (RMSE) for cross-validation (RMSECV), calibration (RMSEC), and testing (RMSET) were calculated using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4.4}$$

where y_i is the measured parameter, $\hat{y}_i$ is the predicted parameter, and n is the number of samples for the data set under consideration. The testing data set consisted of the sixth compact from each design point, whose compaction pressure was pseudo-randomly assigned one of the five possible levels. While this does not constitute a truly independent dataset for model validation (i.e., for use in process control), the course of action is suitable for exploratory analyses such as this.

4.3 Results and Discussion

4.3.1 Optical Effects of Varying Compaction Pressure

The optical effects of varying compaction pressure, which elicits change in the physical parameters of the samples, are difficult to visualize amongst the broad chemical variation built into the design. Thus, Figure 4.1a and Figure 4.1b display the

characteristic baseline shifts for reflectance and transmittance spectra associated with changes in tablet density when the percent contributions of the constituents are unchanging. The baseline slope increases with increasing compaction pressure for reflectance spectra while the opposite trend is observed for transmittance measurements.

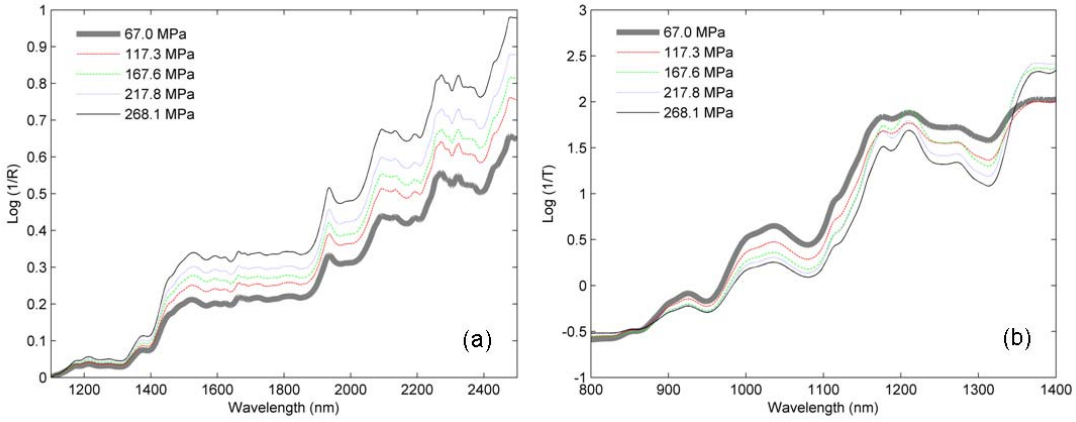

Figure 4.1. Raw NIR absorbance spectra illustrating the spectral effect of compaction pressure when concentration is unchanging for reflectance (a) and transmittance (b) measurements. The design point illustrated is 40% theophylline, 40% lactose, 20% MCC, and 0% starch. The intensity recorded at the first wavelength of each scan was subtracted from all remaining wavelengths to facilitate viewing.

The characteristic increase in measured absorbance as a result of an increase in compaction pressure for NIR reflectance spectra (Figure 4.1a) is consistent with the results published over the last fifteen years.[128-133,137-149] An increase in compaction pressure results in a decrease in reflected intensity. Without the benefit of corresponding transmittance measurements, it is difficult to determine whether the effect is due to specular reflectance, absorption, or transmission of the source radiation. As suggested by Cogdill and Drennen,[124] Figure 4.1b reveals that the latter may well be true; as compaction pressure increases, transmission also increases.

Otsuka *et al*. recently published the results of a detailed scientific study of this effect.[152] The authors cite a light penetration model based on time-of-flight analysis of wood chips, where the transmittance of NIR radiation can be ascribed to the pore structure of the sample. It was shown that as the tablet porosity decreases, the solid/solid boundary surface area increases, resulting in less scattered light and more transmitted light. On the other hand, as the porosity increases, the air/solid boundary surface area increases, and scattering predominates. As for reflectance of NIR radiation, the authors hypothesize that for a tablet having low porosity, the light penetrates deep into the sample and is consequently absorbed between the matrices. Conversely, reflectance intensity is higher in a highly porous tablet due to scattering at the air/solid boundaries, which precludes the radiation from penetrating deep into the sample.[152]

These results dispel a commonly held explanation for NIR reflectance sensitivity in the presence of varying compact density (often referred to as tablet hardness). Therefore, the notion that a more dense compact scatters less and absorbs more may not be entirely accurate. While it is not feasible to generalize these conclusions to all compact systems, it is expected that these results are generally applicable to most compacted pharmaceutical powders. Based on the results of this work, the authors speculate that a reduction in compact porosity or void fraction increases the forward promotion of scattered photons, which is consistent with the phenomena observed in Figure 4.1.

4.3.2 Relative Density Modeling

It was determined that three latent variables (PLS factors) were required to adequately model the effect of relative density variation on reflectance and transmittance

spectra based on the minimization of RMSECV. The use of three latent variables is reasonable considering the number of degrees of freedom in the experimental design. A summary of the calibration statistics is provided in Table 4.3. Calibration and test set prediction plots for both reflectance and transmittance are shown in Figure 4.2. The comparison of predicted and reference relative density values suggest a non-linear relationship; a potential explanation for this is the non-linear relationship between relative density and compaction pressure (Figure 4.3a).

Table 4.3. Calibration statistics for the prediction of relative density and crushing strength for NIR reflectance and transmittance geometries.

Reflectance Geometry		
Data Type	Relative Density	Crushing Strength
Method	SIMPLS	
Preprocessing	Raw Spectra	Raw Spectra
Latent Variables	3	3
RMSECV [unitless, kp]	0.013	4.640
RMSEC [unitless, kp]	0.013	4.182
RMSET [unitless, kp]	0.017	4.739
R^2_{CV}	0.922	0.832
R^2_{CAL}	0.928	0.864
R^2_{TEST}	0.906	0.856
$Bias_{CV}$ [unitless, kp]	1.141E-04	5.301E-02
$Bias_{TEST}$ [unitless, kp]	2.921E-03	-1.437E-01
Transmittance Geometry		
Data Type	Relative Density	Crushing Strength
Method	SIMPLS	
Preprocessing	Raw Spectra	Raw Spectra
Latent Variables	3	5
RMSECV [unitless, kp]	0.018	3.678
RMSEC [unitless, kp]	0.017	3.257
RMSET [unitless, kp]	0.013	3.187
R^2_{CV}	0.850	0.895
R^2_{CAL}	0.868	0.917
R^2_{TEST}	0.936	0.938
$Bias_{CV}$ [unitless, kp]	1.402E-04	9.543E-02
$Bias_{TEST}$ [unitless, kp]	-5.263E-04	-5.638E-01

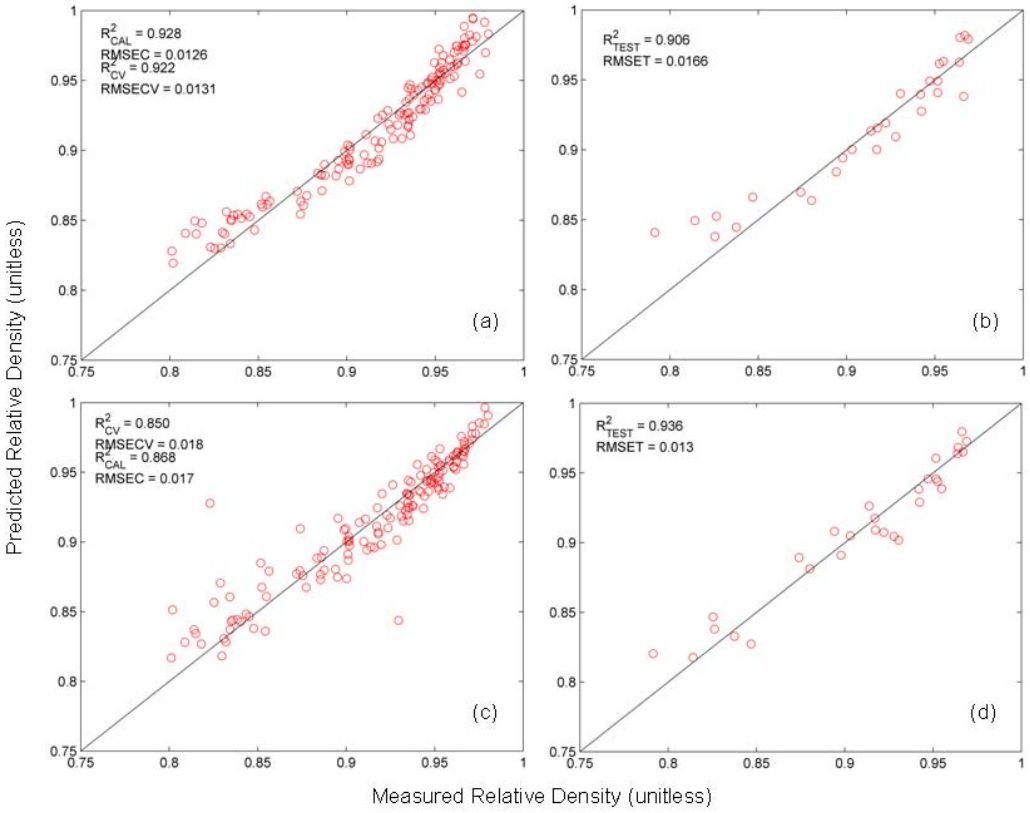

Figure 4.2. Predicted versus measured relative density plots for the calibration (a,c) and test (b,d) data sets for reflectance (a,b) and transmittance (c,d) geometries.

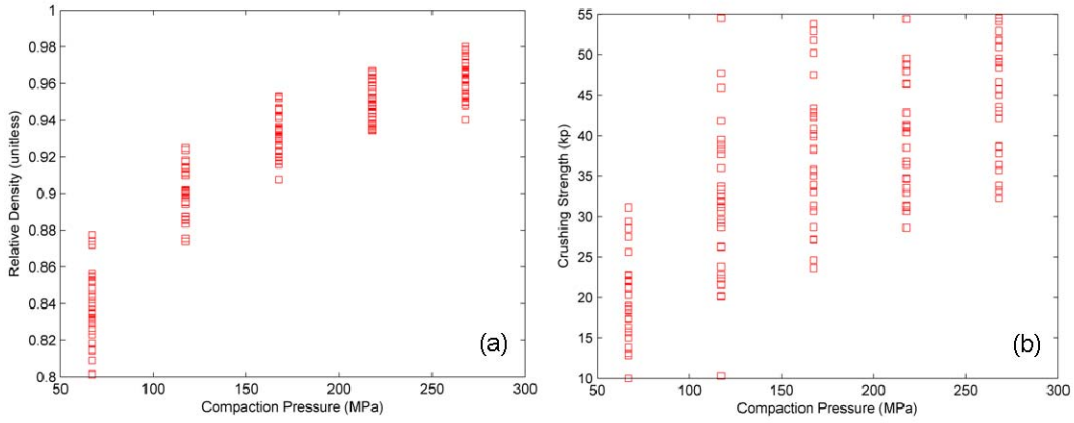

Figure 4.3. Plots of relative density (a) and crushing strength (b) versus compaction pressure. Only calibration samples are plotted.

The PLS loadings, scores, and regression vectors from the optimized calibration models were studied to evaluate the potential link between the physical data modeled and the chemical composition of the compacts. The calibration samples were grouped into different datasets based on constituent concentration and compaction level prior to modeling. Figure 4.4a and Figure 4.4b display the score values on the second versus the first latent variables for transmittance and reflectance geometries, respectively. The majority of the variance (89.07 and 85.69%) associated with the first latent variable is attributable to compaction pressure. This is expected as compaction pressure was one of the independent factors varied in the design; changing compaction pressure results in different relative densities.

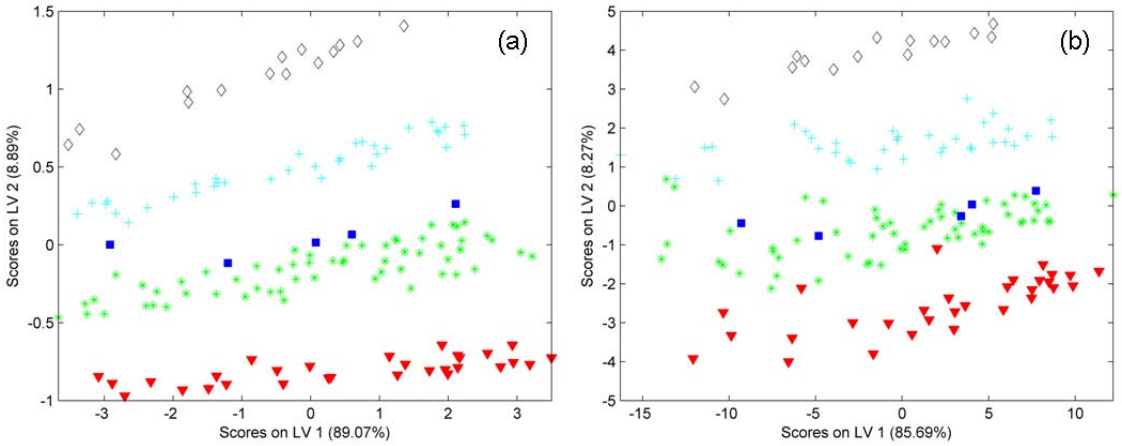

Figure 4.4. Plot of PLS scores for relative density on latent variable two versus latent variable one for reflectance (a) and transmittance (b) where the anhydrous theophylline concentration calibration samples were grouped into the following classes: red-triangle = 0.0 w/w, green-asterisk = 0.2 w/w, blue-square = 0.25 w/w, cyan-plus = 0.4 w/w, and black-diamond = 0.6 w/w.

Similar to differing compaction pressures, changes to constituent concentrations result in varying compressibilities. When the calibration samples were grouped based on theophylline concentration (Figure 4.4), a pronounced separation in score space was

evident along the second latent variable (8.89 and 8.27% variance). This pattern suggests the influence of theophylline on the modeling of relative density as a function of compaction pressure. Its influence was also demonstrated by the second loading that weighted wavelength regions associated with features in the pure component spectrum of anhydrous theophylline (second loading vector not shown). No separation was observed in reflectance or transmittance for the other components among all possible combinations of the first three latent variables.

The PLS regression vectors used for the prediction of relative density appear in Figure 4.5a and Figure 4.5b for reflectance and transmittance geometries, respectively. The regression vectors also have several features correlated to the pure component spectra of anhydrous theophylline, which is not unexpected. A slightly greater correlation was observed between the regression vectors and the pure component spectrum for the reflectance data. The regression vectors further convey the importance of anhydrous theophylline for the prediction of relative density via NIR reflectance and transmittance spectra.

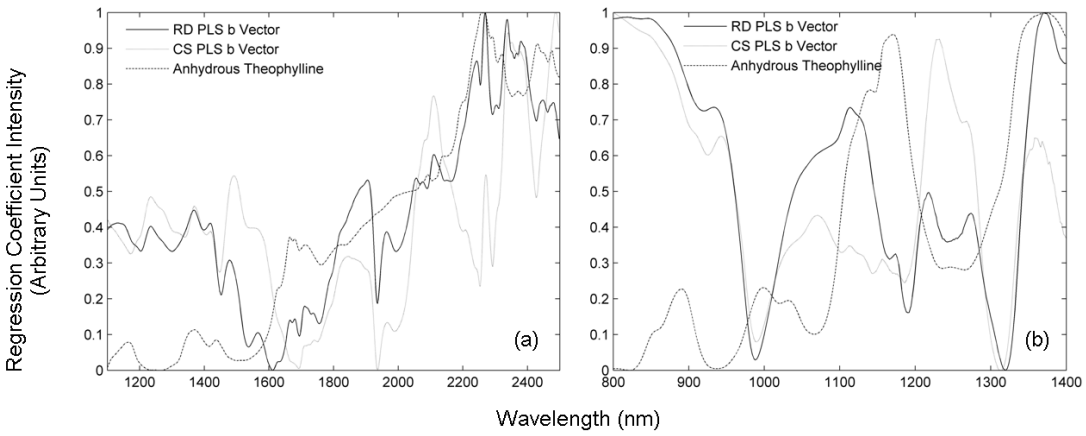

Figure 4.5. PLS regression vectors for relative density (solid) and crushing strength (dotted) and the pure component spectrum of anhydrous theophylline (dashdot) for reflectance (a) and transmittance (b) geometries. All vectors were scaled to the range of 0 to 1 to facilitate viewing.

Pure component compacts were produced to assess which component (1) produced compacts having the highest relative densities at a given compaction pressure and (2) was most susceptible to changes in relative density upon compaction. The pure component compaction profiles evaluating relative density versus compaction pressure for all four constituents after production are shown in Figure 4.6. The trend did not change following the approximately 55 day relaxation period (plot not shown). Of the four components, anhydrous theophylline consistently produced pure component compacts with the highest values. Relative density can also be used to estimate the porosity (ε) of a cylindrical compact from the equation

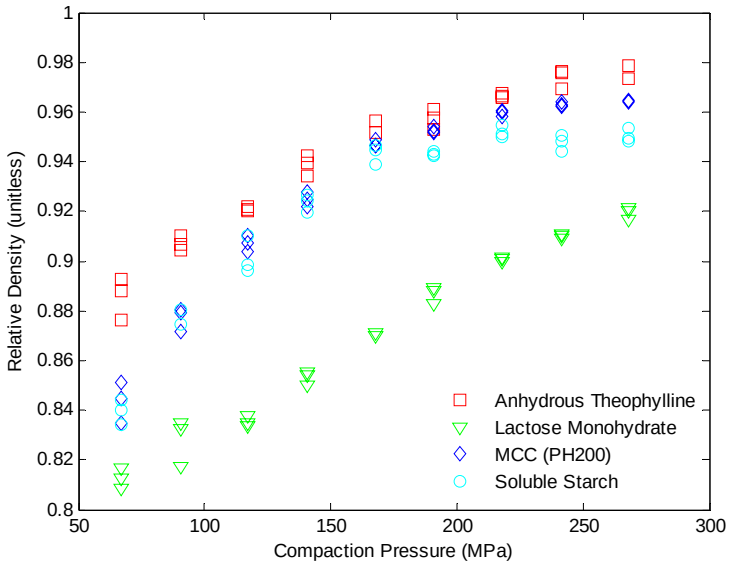

Figure 4.6. Relative density versus compaction pressure profiles for anhydrous theophylline (squares), lactose monohydrate (triangles), MCC (diamonds), and soluble starch (circles) immediately after production.

$$\varepsilon = 1 - \frac{\rho_{comp}}{\rho_{true}} \tag{4.5}$$

where ρ_{comp} and ρ_{true} are the same as previously defined. Applying Equation 4.5 to consider the results shown in Figure 4.6 in terms of porosity (void fraction), anhydrous theophylline consistently produced the least porous pure component compacts. As far as the gross change in porosity is concerned, MCC displayed the largest change in porosity over the compaction range tested. This is not unexpected considering the microstructure and propensity of MCC to deform under mechanical stress.[153,154] It was expected that MCC would have a large influence (leverage) on the spectroscopic analyses; however, this affect was not observed. This may be explained by the spectroscopic signals of each constituent.

The pure component spectrum of anhydrous theophylline is more orthogonal to the interfering factors; there was much greater collinearity between lactose monohydrate, MCC, and soluble starch (Figure 4.7). This orthogonality enhances the net analyte signal for anhydrous theophylline, which inherently augments the sensitivity of the method to this material. Therefore, the wide anhydrous theophylline concentration range tested (Table 4.1) most likely overwhelmed the signal attributable to the other three constituents.

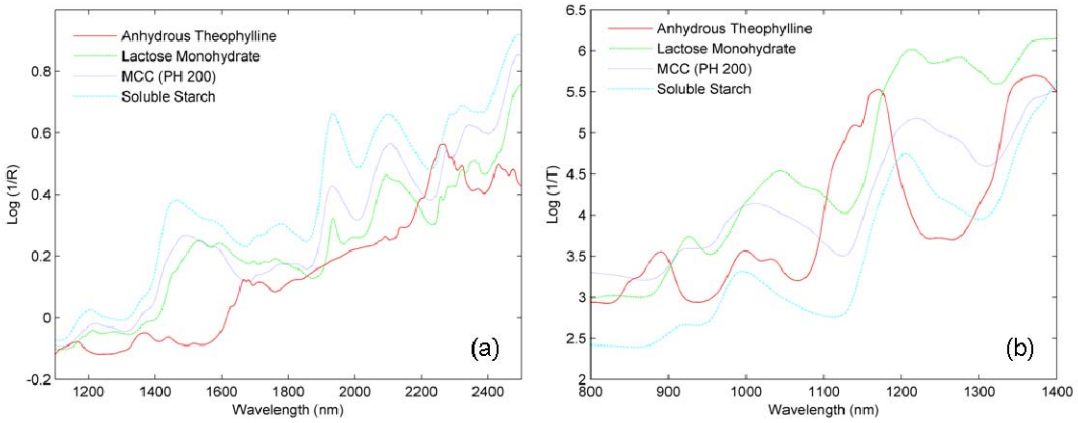

Figure 4.7. Pure component spectra for transmittance (a) and reflectance (b) geometries. Individual vectors appear as the mean spectrum of twenty-seven pure-component compacts varying over the compaction pressure range of 67.0 - 268.1 MPa.

4.3.3 Crushing Strength Modeling

Based on the minimization of RMSECV, it was determined that three latent variables were required to adequately model the crushing strength variation for reflectance data while five were required for transmittance measurements; this result was independent from relative density modeling. A summary of the calibration statistics is provided in Table 4.3. These data demonstrate that there was sufficient sensitivity in both the NIR reflectance and transmittance data to model crushing strength as observed

in the relative density analyses. This can be observed in the predicted versus measured crushing strength plots for the calibration and testing data sets (Figure 4.8). It should be noted that the inclusion of the samples reading 55.9 kp would not have considerably altered the results shown in Figure 4.8 and Table 4.3. These data were excluded as they were at the limit of the testing instrumentation.

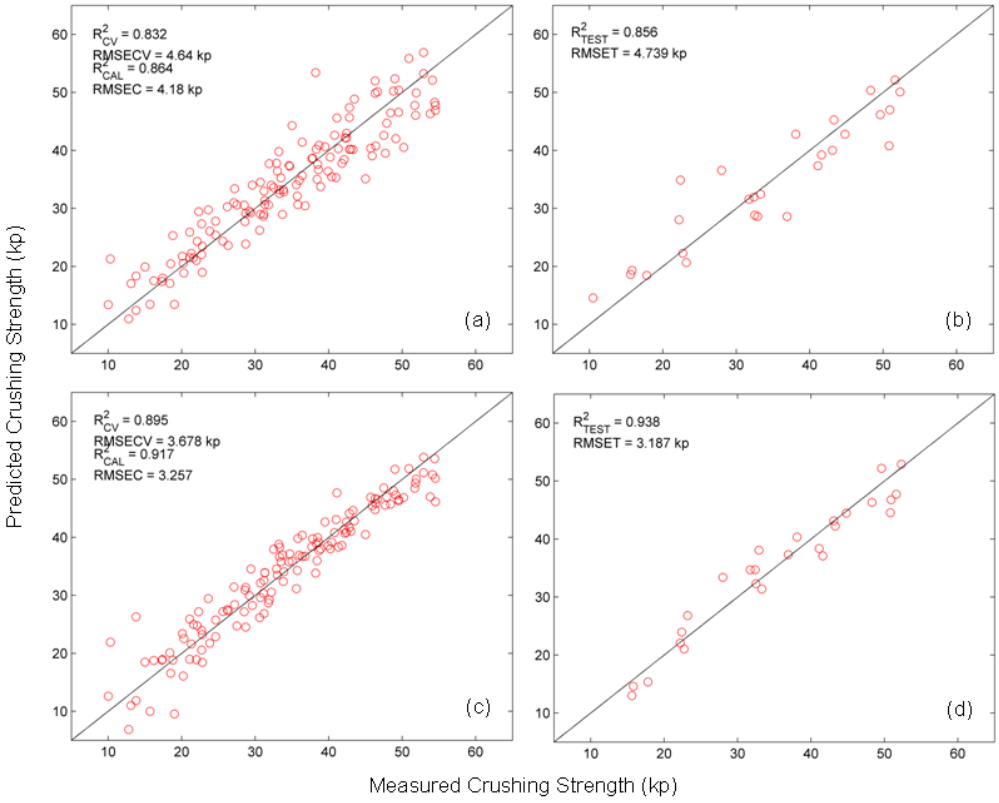

Figure 4.8. Predicted versus measured crushing strength plots for the calibration (a,c) and test (b,d) data sets for reflectance (a,b) and transmittance (c,d) geometries.

Comparing the plots in Figure 4.8 with those in Figure 4.2, it is apparent that the predicted versus measured plots for crushing strength demonstrate a more linear relationship than was observed for relative density. However, there is an apparent reduction in model accuracy, which may be attributed to the reduced precision of the

reference measurements. The crushing strength calibration may be more linear than the relative density model because the measurement is more closely analogous to the structural factors affecting photon propagation, namely particle-particle interaction. If the tensile strength of an object at zero porosity is defined as the intrinsic strength of the material(s) multiplied by the area normal to the applied force, then greater compaction pressures must increase the effective area of particle-to-particle contact, even as compact volume is reduced. As more and more particles fuse together, the effective path-length of photons through the material is enhanced, decreasing the reflectance and increasing the transmittance intensities of the samples. It is important to note that an increase in particle fusion will also enhance (to a certain extent, although the enhancement is ultimately limited by the material's crystallographic density) the relative density of a sample, having an analogous effect on the reflectance and transmittance intensities.

Subsequent to model optimization, the PLS scores, loadings, and regression vectors were analyzed for a relationship between physical and chemical information within the spectra. In a similar manner, the calibration samples were grouped according to compaction and concentration levels. Separation of the samples by compaction pressure demonstrated that the variance associated with the first latent variable is most likely due to consolidation. The separation among different compaction levels was not as clear for the crushing strength analysis. This may be explained by the relatively lower precision of diametric compression testing; note the tighter distribution of data points around the unity line for relative density (Figure 4.2 and Figure 4.8). This uncertainty effectively reduces the covariance between the NIR spectra and the reference data, which decreases the likelihood of separation in scores space. Ultimately, this reduction in

covariance deflates the signal-to-noise ratio, causing the data for adjacent compaction levels to blur together, perhaps because of increased rotational ambiguity during solution of the factor structure.

As was previously observed in the relative density analyses, a division in score space is evident when the calibration samples are grouped according to anhydrous theophylline concentration for both reflectance and transmittance data (Figure 4.9a and Figure 4.9b, respectively). This separation indicates the influence of anhydrous theophylline (on the calibration model) for the prediction of crushing strength. The influence of anhydrous theophylline is further demonstrated by the second loading vector that tracked the loss of theophylline (observed for both sensing geometries). Several segments of the vectors corresponded to features in the pure component spectrum of anhydrous theophylline (loading vector not shown). No discernible patterns were observed for the remaining components among all possible combinations of the first three and five latent variables for reflectance and transmittance geometries, respectively.

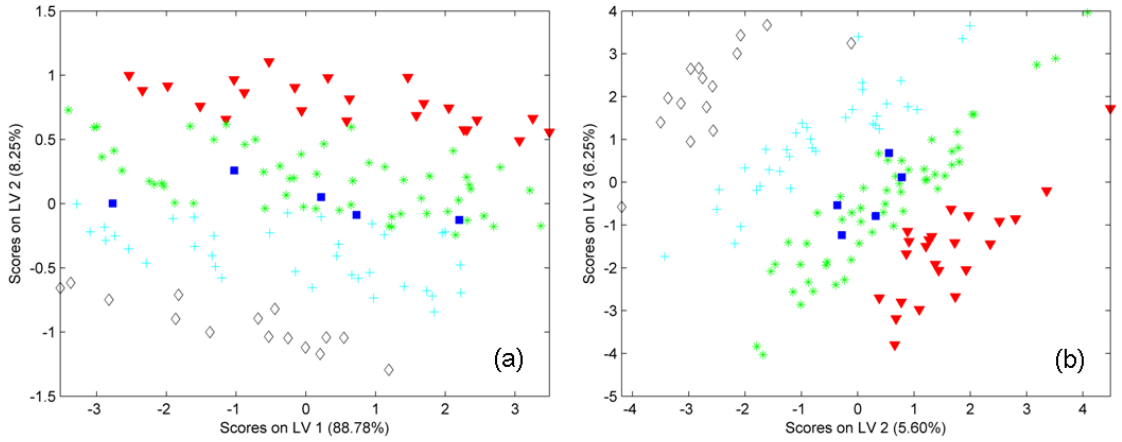

Figure 4.9. Plot of PLS scores for crushing strength on latent variable two versus latent variable one for reflectance (a) and latent variable three versus latent variable two for transmittance (b) where the calibration samples were grouped into classes (see Figure 4.4 caption) based on anhydrous theophylline concentration.

The crushing strength regression vectors have several features related to the pure component spectrum of anhydrous theophylline (as observed in the relative density analyses). The PLS regression vectors used for the prediction of crushing strength appear in Figure 4.5. Once again, the correlations between the regression vectors and anhydrous theophylline were slightly less intense for the transmittance data.

While models employing baseline fitting algorithms have proven effective in the presence of both chemical and physical variation,[130] PLS regression was used due to the extended concentration range (0 – 60% w/w for all constituents). When composition was held constant (by selecting sub-sets of the calibration data), baseline fitting methods performed comparably with PLS regression (results not shown). Additionally, this regression method was implemented based on its tangible outputs (e.g., scores, loadings, and full-spectrum regression vectors), which can be used to better understand the effects of chemical variation on physical parameters of pharmaceutical compacts.

To assess the effect of concentration variation, 29 sub-calibrations where constructed using linear regression to relate the standard deviation of each spectrum to relative density or crushing strength. In order to mitigate concentration's leverage, a sub-calibration was built at each experimental design point (Table 4.1); only calibration samples corresponding to the above models were used. For the prediction of relative density, the mean, median, and standard deviation of the sub-calibrations' coefficient of determinations were (0.954, 0.957, 0.027, n = 29) and (0.529, 0.578, 0.333, n = 29) for reflectance and transmittance geometries, respectively. The ability to model relative density as a function of the standard deviation of the spectrum is compromised in the presence of the extended concentration range as evidenced by the reduction in R^2 when

all 145 calibration samples are assessed simultaneously (0.840 and 0.189 for reflectance and transmittance geometries, respectively).

The ability to predict crushing strength as a function of the standard deviation of each spectrum is compromised in the presence of concentration variation when all 133 calibration samples are modeled (R^2 = 0.511 and 0.005 for reflectance and transmittance geometries, respectively). For the prediction of crushing strength, the mean, median, and standard deviation of the sub-calibrations' coefficient of determinations were (0.945, 0.970, 0.079, n = 29) and (0.436, 0.337, 0.358, n = 29) for reflectance and transmittance geometries, respectively.

When concentration is held constant, both relative density and crushing strength are modeled well by the standard deviation of the log (1/R) spectra. The authors suggest that the thickness of the compacts (mean = 4.387 mm, median = 4.307 mm, standard deviation = 0.237 mm, n = 174) in conjunction with tablet positioning error[155] may have compromised the integrity of the transmittance results due to nonlinearity.

4.3.4 Rationalization of NIR Sensitivity to Compact Density

It is proposed that NIR sensitivity to variation in compact density is attributed to changes in the distribution of forward and back photon propagation. It is well understood that scattering events occur only where there is a transition in refractive index.[156] In the absence of a transition, an impinging photon will tend to be transmitted without a change of direction. Similarly, when the transition is small (e.g., between particles of different materials having similar refractive indices), the change in photon direction will be small. Particle-particle interfaces present a smaller refractive index change relative to air-particle interfaces.

When contrasting the photon path in highly porous and highly dense compacts, the latter present fewer air-particle interfaces, and more particle-particle interfaces. Consequently, photons incident on less porous compacts encounter fewer scattering opportunities per unit of pathlength, and tend to exhibit enhanced forward propagation due to increased particle-particle transmission. It is understood that individual constituents have different optical properties, and therefore, each would be expected to exhibit a unique optical response to varying porosity. This argument is similar to that posed by Isaksson, Miller, and Naes, where they investigated the spectral effects of laminate films covering homogenized meat.[157] The authors found that the addition of the laminate, whose refractive index is more similar to that of the meat than the air, reduced the specular scattering observed at the laminate-meat interface (as opposed to the air-meat interface). Ultimately, the addition of the laminate film decreased the diffuse reflectance intensity and increased the transmitted light intensity.

As noted above, an increase in compaction pressure (alternatively, relative density and crushing strength) resulted in higher amounts of transmitted radiation. For the compacts studied, increasing density did not lead to higher absorbance intensities as has been previously asserted.[128,130-132,141] While the data necessary to definitively state the reason(s) why this occurs have yet to be published, the theory of large-particle (e.g., Mie) scattering may offer a plausible first approximation.[156,158]

For large-particle scattering, intensity can be viewed in terms of its forward and back distribution (or transmission and remission), where for particles larger than wavelength of the impinging radiation, interference is such that the forward direction is favored.[156,158] If we can assume that, as porosity decreases, the average size of scattering

centers (e.g., particle size) within the compact increases, Mie theory would predict a greater propensity for scatter in the forward direction. Moreover, as the void fraction decreases, the relative area of low refractive index (particle-particle) transitions in comparison to high refractive index (air-particle) transitions increases, resulting in scattering (refraction) angles favoring forward propagation.[156,158] Forward scatter does not necessarily favor additional scatter more so than absorption; however, the reduction of scattering in the reverse direction would ostensibly reduce the reflectance intensity while increasing the transmittance intensity. Moreover, this would lead to increased pathlength for diffuse reflectance measurements, and reduced pathlength for (diffuse) transmittance measurements. The specific application of Mie theory to consolidated pharmaceutical powders, however, violates many of the underlying assumptions (e.g., isolated particles, spherical particles, single scattering events). [156,158]

Taking into consideration the anisotropy of both pharmaceutical solid materials and the composites formed by their compaction, it is reasonable to consider alternative methods which are valid for particles having non-spherical shape and contact with other particles. Grundy, Douté, and Schmitt proposed a model based on ray tracing and Monte Carlo simulation of the linear polarization and scattering by anisotropic particles of sizes much larger than the wavelength of incident light.[159] While this reference dealt exclusively with individual particles, their S-Scat model yielded results similar to those predicted by Mie theory. Additionally, the authors observed that for irregular particles, scattering in the forward direction (larger phase angles) was favored.[159] Utilizing models with the ability to compensate for the anisotropy of pharmaceutical solids (e.g., S-Scat)

may help to clarify the spectroscopic intricacies observed during the modeling of relative density and crushing strength.

Additional research is required to address concomitant issues such as variation in raw material particle shape, size, etc. to understand how photons interact with compacts of varying density and composition. Future research will be focused on developing comprehensive NIR photon migration simulation algorithms to address these issues.

4.4 Conclusions

Calibration models were constructed using PLS regression to relate NIR reflectance and transmittance spectra to compact relative density and crushing strength values. Both reflectance and transmittance data had adequate sensitivity to model both physical characteristics. Second to compaction pressure, anhydrous theophylline was most influential on the spectral analyses of relative density and crushing strength, which was mainly a function of the experimental design and the enhanced chemical sensitivity to this particular component. Additionally, the calibration models for relative density and crushing strength demonstrated that the latter was more linear but less accurate in prediction. This study also demonstrated that, in contrast to the existing interpretations, increasing tablet density does not necessarily reduce scattering or increase absorbance of NIR radiation. The optical interactions are a function of changing compact porosity, which promotes greater scattering in the forward direction. Propagation of light in the forward direction is detected as an increase in transmitted intensity. Future studies will focus on using empirical data combined with *in silico* simulations to discern what happens to source photons as they interact with compacted pharmaceutical powders of varying relative densities.

Chapter 5: Determination of Figures of Merit for Near-Infrared and Raman Spectrometry by Net Analyte Signal Analysis for a Four Component Solid Dosage System

The material presented in Chapter 5 was previously published in Short SM, Cogdill RP, Anderson CA 2007. Determination of figures of merit for near-infrared and raman spectrometry by net analyte signal analysis for a 4-component solid dosage system. AAPS PharmSciTech 8(4):Article 96. DOI: 10.1208/pt0804096.

5.1 Introduction

A number of technologies may be suitable for a given analytical measurement application. Ultimately, a decision must be made as to which device will be deployed. Common methods for comparison of instruments based upon different fundamental principles are not well established. While there are multiple considerations upon which instrument selection is based (cost, performance, infrastructure, etc.), this work will focus only on method performance characterization for two sample technologies.

The advent of the Process Analytical Technology (PAT) initiative[6] has increased the performance demands upon, and the need for understanding of, analytical methods. In a PAT environment, sensors are controlling processes (as opposed to advising). Pharmaceutical literature often gives the impression that a multivariate sensor is generally applicable for in-, on-, and at-line process monitoring applications. These applications typically require chemometric modeling/calibration[160] to transform instrument signal into useable information. Multivariate calibration provides a platform for data driven analyzer selection.

Process analytic measurement applications in pharmaceutical science include qualitative and quantitative identification of compounds, examination of phase transformations and polymorphs, and investigations of manufacturing and process development. Near-infrared (NIR) and Raman spectroscopy are both sensitive to, and capable of accurately predicting, these phenomena.[124,161] As NIR is based on optical absorption and Raman on the measurement of inelastic scattering, comparison of the parallel performance of these two classes of instrumentation is indirect. More often than not, the instrument eventually deployed is selected based on limited criteria which do not consider all aspects of performance.

Calibrations are frequently evaluated using the coefficient of determination (R^2) and/or estimation of prediction error. [162-166] Neither of these measures directly consider issues such as precision[164,165] and signal-to-noise (S/N) ratio.[167,168] While these are important statistics, they should not be considered independently; rather, the effectiveness of a method should be judged based on a complete assemblage of indicators which describe all aspects of performance and are generalized across technological platforms. Net analyte signal (NAS) theory provides a mechanism for determining figures of merit indicative of an instrument's utility.

Net analyte signal theory is the concept of separating relevant signals for a particular component of interest from the remaining interfering elements present within the spectra.[169] Lorber[170] is widely acknowledged as the originator of multivariate NAS theory; Brown,[171] however, makes the clarification that Morgan[172] published on a similar topic prior to Lorber, although the work "contains some errors."[171] Net analyte signal provides a tool for calculating multivariate figures of merit; prior to Morgan/Lorber,

techniques similar to NAS were applied only to data from univariate methods.[173] A detailed description of univariate calibration and the corresponding determination of figures of merit are outside the scope of this work. For a more detailed discussion, please refer to Olivieri *et al*. and their references.[173]

Multivariate NAS was first implemented using pure component projection[170] and classical regression;[174] *i.e.*, constituent concentration was directly related to instrumental response vectors. However, implementation of NAS via pure component or classical regression methods is cumbersome in that the spectra or concentrations for all contributing species are required.[169] The implementation of multivariate NAS was solved in terms of inverse regression by Lorber, Faber, and Kowalski, requiring only the concentrations for the component(s) of interest to be known.[169] For inverse regression, concentration is expressed as a function of instrumental response.[174] Nonetheless, both classical and inverse regression mathematics are suitable for the determination of NAS; however, the remaining descriptions and the work herein strictly assume the application of inverse regression.

Mathematically, NAS is defined as the portion of signal unique to the constituent being considered and is orthogonal to all other factors present in the data.[169,170] NAS, therefore, comprises the signal directly useful for quantification.[169] A mixture spectrum (**r**) extracted from a spectral matrix containing multiple constituents can be resolved into

$$\mathbf{r} = \mathbf{r}^{*} + \mathbf{r}^{\perp} \tag{5.1}$$

where $\mathbf{r}^{*}$ and $\mathbf{r}^{\perp}$ are mutually orthogonal components representing the NAS vector for the particular component of interest and the vector of interferences, respectively.[169] However, it is well understood that controllable and uncontrollable errors influence the

performance of any analytical technique. This error (ε), which will be slightly different for each sample acquired, can also be partitioned into its respective mutually orthogonal components: the part that is orthogonal to the interferences (ε^*), and the portion that lies within the interference space ($\varepsilon^{\perp}$).[175] It is important to understand that the estimated $\mathbf{r}^*$ will not lie exactly in the true direction of the NAS vector as a result of ε^*.[175] The former (ε^*) is the portion of total stochastic error (ε) which contributes to imprecision.

Multiple algorithms for calculating NAS have been reported in the literature,[176,177] with the method of computation and resulting output differing. In this work, the method proposed by Bro and Andersen is utilized.[178] All equations use **X** to represent spectral matrices; bold characters other than **X**, *e.g.*, **x**, **y**, and **NAS**, represent vectors; small, italicized characters, *e.g.*, *x* and *y*, represent scalars. Additionally, the notation ||**x**|| signifies the Euclidean norm of a vector, the square root of the sum of the squared elements. The superscript *T* indicates the transpose of a vector.

Bro and Andersen calculate the net analyte signal vector for a particular component of interest by

$$\mathbf{N\hat{A}S}_i = \left(\mathbf{x}_i \cdot \mathbf{b}\right) \cdot \left(\mathbf{b}^T \cdot \mathbf{b}\right)^{-1} \cdot \mathbf{b}^T \qquad (5.2)$$

where $\mathbf{x}_i$ is a sample spectrum from matrix **X** and **b** is a column vector of the regression coefficients for **X**;[178] principal components regression (PCR) or partial least-squares (PLS) are common regression techniques used to estimate **b**.[160] It should be noted that **X** is corrected for the mean; thus, outputs from computations employing $\mathbf{N\hat{A}S}$ are mean-centered. Results can be rescaled using the vector of means from the centering operation of **X** to the original range. Net analyte signal can also be expressed in scalar form, with no loss in information (but reduced interpretability), by the equation[169]

$$\hat{NAS}_i = \left\| \mathbf{N}\hat{\mathbf{AS}}_i \right\| \tag{5.3}$$

As Lorber *et al*. have noted, the term "NAS" is often interchangeably used to represent the NAS vector and the scalar NAS quantity.[169] Additional discussion concerning the mathematics behind the determination of net analyte signal can be found elsewhere.[169-172,175-181]

5.2 Materials and Methods

5.2.1 Tablet Production

A fully-balanced, four-constituent mixture design comprised of anhydrous theophylline (Lot No. 92577, Knoll AG, Ludwigshafen, Germany), Lactose 316 Fast Flo NF Monohydrate (Lot No. 8502113061, Hansen Labs, New Berlin, WI), microcrystalline cellulose (MCC, Avicel PH 200, Lot No. M427C, FMC BioPolymer, Mechanicsburgh, PA), and soluble starch GR (Lot No. 39362, EMD Chemicals Inc., Gibbstown, NJ) was generated. The approximate median particle size of the theophylline, lactose, MCC, and starch (reported by documentation from their respective suppliers), was ~90, ~100, ~180, and ~17 microns, respectively. No further analysis or alterations were performed on the materials to determine or alter the particle size distribution. Twenty-nine design points were chosen to cover a wide range in all constituents and to remove any possibility of factor aliasing (Figure 5.1). Analysis of the mixture covariance matrix (not shown) demonstrated the design is balanced in all directions, giving equal emphasis to all constituents.

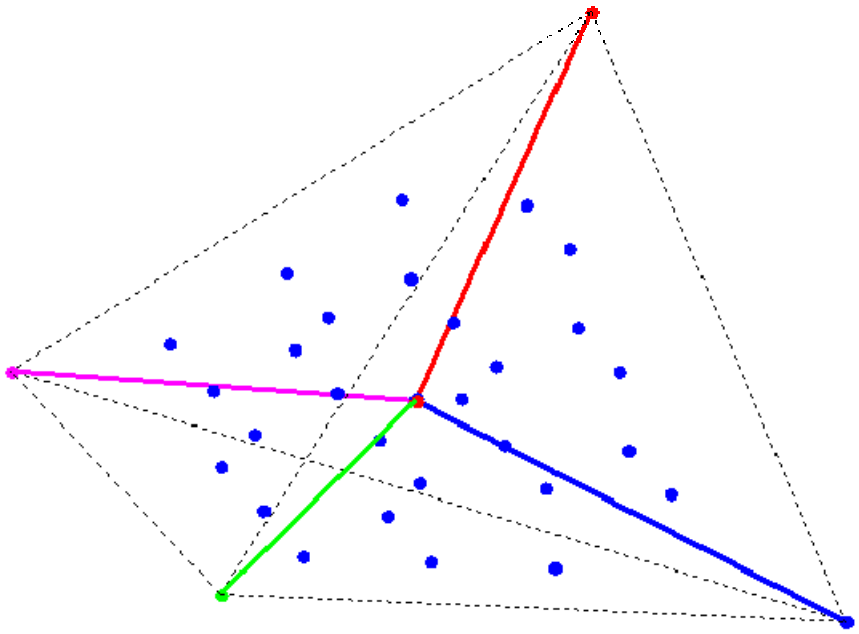

Figure 5.1. Diagram illustrating the concentration design matrix. The four-component design matrix can be viewed in three dimensions, and is in the shape of a tetrahedron because of the constraint imposed by concentration closure. The center of the pyramid represents the point of equal concentration in all components and the vertices represent the points of pure constituents. The solid lines are not labeled for their corresponding constituents as they are all equivalent due to the balanced design.

Materials for each design point mixture were dispensed by weight (Data Range, Model No. AX504DR, Mettler Toledo, Columbus, OH) in accordance with the design, and were immediately transferred to 25 mL glass scintillation vials. In total, 6000 mg of material was weighed out for each point, and the nominal weights for all constituents were adjusted to the observed mass data to calculate actual concentration. After all components were added to each vial, they were tumbled for 5 minute cycles on a rotating Jar Mill (US Stoneware, East Palestine, OH, USA). After every blending period, each vial was manually inverted to collect a NIR reflectance spectrum directly through the bottom of the glass (FOSS NIRSystems 5000, FOSS NIRSystems, Inc., Laurel, MD). An *ad hoc* partial least-squares II (PLS-2) calibration, using the constituent concentrations as reference data, was constructed after each blending cycle to assess homogeneity.

Mixtures were assumed to be homogeneous when further mixing failed to yield an increase in the calibration's coefficient of determination.

The mixtures from each design point were then subdivided and tableted at 5 levels of compaction force (67.0, 117.3, 167.6, 217.8, 268.1 MPa) on a Carver Automatic Tablet Press (Model 3887.1SD0A00, Wabash, IN) using flat-faced punches and a 13 mm die. A dwell time of 10 seconds was employed. Six compacts were produced per design point, with the sixth tablet's compaction force randomly selected from one of the five possible levels. The compaction order was randomized to ensure homoscedasticity of experimental error. In total, 174 compacts were produced with a nominal target weight of 800 mg per tablet. A small arrow was drawn on the perimeter of each tablet (to avoid spectroscopic interference) to distinguish between the two flat surfaces. Tablet preparation and compaction occurred over a three day period, after which the compacts were left to relax for 15 days prior to spectroscopic analysis to compensate for any radial and/or axial tablet expansion.

5.2.2 Data Acquisition, Instrumentation, and Software

Near-infrared reflectance measurements (expressed as log(1/R), or absorbance intensity) for both sides of each tablet were acquired over the wavelength range of 1100 – 2498 nm at a 2 nm increment (FOSS NIR Systems 5000, FOSS NIRSystems, Inc., Laurel, MD). Thirty-two sub-scans were accumulated for each resultant sample spectrum. Prior to scanning, the tablets were precisely centered using the positioning iris standard on this particular instrument. Absorbance data were collected by way of Vision data acquisition software (version 2.00, FOSS NIRSystems Inc.) and exported in .NSAS file format. Two *ad hoc* partial least-squares II (PLS-2) calibrations, using the

constituent concentrations as reference data, were constructed from spectra corresponding to a specific surface of the tablets. Since the coefficients of determination did not differ until the third decimal place, it was decided to only consider measurements of one tablet face (arrow pointing upwards) for both technologies.

Raman data were measured using a prototype *P^{h}AT System* spectrometer with a laser excitation wavelength of 785nm and equipped with a fiber-coupled probe head (Kaiser Optical Systems, Inc., Ann Arbor, MI). The *P^{h}AT System* samples a spot size of approximately 6 mm. Two accumulations were acquired per scan employing an integration time of 10 seconds over the range of -64.2 to 1895.7 cm^{-1} at a 0.3 cm^{-1} increment. A dark scan was subtracted and the cosmic ray filter and intensity calibration options were selected. The tablets used in this study were larger than could be accommodated in the tablet holder located in the sample chamber of the *P^{h}AT System* requiring the tablets to be manually positioned such that the laser spot was visually centered on the flat face. Raman intensity data were collected via the HoloGRAMS software package (version 4.0, Kaiser Optical Systems, Inc.) and exported in .GRAMS file format.

The Unscrambler (version 9.0, Camo Software Inc., Woodbridge, NJ) was used to convert both .NSAS and .GRAMS files to .mat format to allow further data manipulation. All spectral data were analyzed in the Matlab environment (version 7.1, The MathWorks, Natick, MA) using the PLS_Toolbox (version 3.0, Eigenvector Research, Inc., Manson, WA) and software developed by the Duquesne University Center for Pharmaceutical Technology (DCPT).

5.2.3 Partial Least-Squares Analysis

NIR and Raman data were analyzed separately but in an identical fashion. The NIR spectral range and resolution were not altered; however, the Raman spectral range was truncated to 205.5 - 1895.7 cm^{-1} to remove residual Rayleigh line radiation and to reflect the operating range of the analyzer. Prior to calibration, the Raman data were evaluated using a moving-window calibration technique[179] with various window widths to determine if further wavenumber truncation would be beneficial. Wavelength selection was not shown to enhance calibration performance. Partial least-squares regression[122] was used via the SIMPLS algorithm[123] to relate spectroscopic response to concentration for each constituent on an individual basis. Since analyte concentration is incorporated in the denominator of some figure of merit calculations, samples having a corresponding zero concentration for the component being considered were excluded. Therefore, samples included in the actual calibration data sets were unique for each component.

Preprocessing routines, including standard normal variate (SNV) scaling, detrending, derivatives, and combinations of the preceding were tested.[124] The most favorable data pretreatment method was selected based on a minimization of "batch-wise" cross-validation error, where the batches in this instance were the five different compaction levels. During each cross-validation iteration tablets produced at a particular compaction force were removed, a calibration was constructed, and the concentrations of the excluded samples were predicted via the temporary model. This procedure was continued until all of the samples had been predicted, thereby allowing for the

determination of the root-mean-standard error (RMSE). The RMSE for cross-validation (RMSECV) and calibration (RMSEC) were calculated using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{5.4}$$

where y_i is the measured concentration, $\hat{y}_i$ is the predicted concentration, and n is the number of samples for the data set under consideration. According to ICH guidelines, accuracy expresses the agreeability between reference and predicted values.[182] Therefore, RMSEC and RMSECV are used to report the accuracy of the selected calibration.

5.2.4 Multivariate Figures of Merit

The optimal number of latent variables selected during the estimation of each PLS regression vector was determined by minimizing RMSECV and RMSEC. Once established, the regression vector was used to determine the NAS according to equations (5.2) and (5.3). Given the number of chemical constituents and physical factors varying in this design, it was anticipated that no more than 4 latent variables would be required; however, models with greater rank are feasible but would be increasingly difficult to justify. The NAS vector affords the opportunity to calculate numerous figures of merit, such as sensitivity, analytical sensitivity, selectivity, and S/N ratio. Figures of merit can be determined for every sample using the following formulae.

Sensitivity characterizes the extent of signal variation as a function of analyte concentration; the higher the sensitivity, the greater the instrumental response to an increase in concentration.[169,170] Sensitivity is calculated as[169,173,177]

$$\hat{\mathbf{SEN}}_i = \frac{\hat{\mathbf{NAS}}_i}{y_i} \tag{5.5}$$

where $\hat{\mathbf{SEN}}_i$, $\hat{\mathbf{NAS}}_i$, and y_i are the vector of sensitivities for each instrument variable, the net analyte signal vector, and the measured concentration for the i^{th} sample, respectively. Sensitivity is reported in units of instrument intensity per concentration. Measured concentrations are autoscaled before being utilized in equation (5.5). It is also possible to express sensitivity as a univariate figure of merit by taking the Euclidean norm of the sensitivity vector

$$\hat{SEN}_i = \| \hat{\mathbf{SEN}}_i \| \tag{5.6}$$

where $\hat{SEN}_i$ is the univariate measure of sensitivity for the i^{th} sample.[177,180] Sensitivity is reported in this document as the mean of the univariate sensitivity values for all samples under consideration.

Sensitivity is only applicable to calibrations constructed on devices operating under the same fundamental principles because it incorporates units of instrument signal (NIR absorbance intensity and Raman scattering intensity). The parameter analytical sensitivity (γ) was developed to provide an impartial assessment between dissimilar analytical techniques.[180] Analytical sensitivity is calculated as

$$\gamma = \frac{\hat{SEN}}{\delta r} \tag{5.7}$$

where $\hat{SEN}$ is the mean of the sensitivity values under consideration found using equation (5.6) and δr is a measure of instrumental noise. This normalization procedure allows direct comparison of the measure of sensitivity associated with NIR and Raman

data. Analytical sensitivity has the dimensions concentration^{-1}. δr was determined as the mean standard deviation of $\hat{NAS}_i$ (for the component under consideration) of four tablets broadly varying in constituent concentrations. Additionally, this figure of merit allows an estimation of the minimum discernible concentration difference given the dynamic range modeled[180] (γ^{-1}); this is referred to as *effective resolution*.

Predicted values were determined by[183]

$$\hat{\mathbf{y}} = \mathbf{X} \cdot \mathbf{b} \tag{5.8}$$

where **X** is the spectral matrix and **b** is the PLS regression vector, which varies depending on the number of latent variables applied. It should be noted that the concentration data have been previously autoscaled; thus, predicted values need to be rescaled using the mean and standard deviation of the measured concentrations before the accuracy is determined via equation (5.4).

It is important to observe that all predicted values herein are independent of any NAS calculations performed. Certain NAS techniques allow the determination of predicted concentrations using the equation[169,177]

$$\hat{y}_i = \frac{\| \hat{\mathbf{NAS}}_i \|}{\| \hat{\mathbf{SEN}}_i \|} \tag{5.9}$$

where $\hat{\mathbf{NAS}}_i$, $\hat{\mathbf{SEN}}_i$, and $\hat{y}_i$ are as defined previously. Considering equation (5.5) for the calculation of the sensitivity vector, the NAS method employed within this paper forbids the use of equation (5.9) because it forces $\hat{y}_i = y_i$.

Selectivity is a dimensionless univariate measure of the portion of instrumental signal that is not lost due to spectral overlap, in other words the quantity of signal

unaffected by interfering factors, restricted to a value between 0 and 1.[169] This statistic is calculated for each sample by[169,181]

$$\hat{SEL}_i = \frac{\| \hat{\mathbf{NAS}}_i \|}{\| \mathbf{x}_i \|} \tag{5.10}$$

where $\hat{\mathbf{NAS}}_i$ and $\mathbf{x}_i$ are the NAS vector and the original spectrum for the i^{th} sample. The magnitude of the selectivity parameter is directly dependent on the degree of spectral interference associated with the particular analyte under consideration. Selectivity is reported in this document as the mean of the selectivity values for all samples under consideration.

Signal-to-noise ratio is one of the most important metrics for general comparison of methods. It is calculated as[169]

$$\hat{S/N}_i = \frac{\hat{NAS}_i}{\delta r} \tag{5.11}$$

where $\hat{NAS}_i$ is the scalar representation of the NAS vector, and δr was estimated as the mean standard deviation of the predicted concentrations (for the component under consideration) of four tablets broadly varying in constituent concentrations (i.e., the same four tablets previously used). Linear regression was performed between measured concentration and the univariate NAS values in order to estimate scale (a_1) and offset (a_o) coefficients to transform the NAS value to units of concentration. This enables S/N to be a dimensionless statistic for this four constituent mixture design.

$$\hat{S/N}_i = \frac{a_1 \cdot \left(\hat{NAS}_i \right) + a_o}{\delta r} \tag{5.11a}$$

Signal-to-noise ratio is reported in this document as the mean of the S/N values for all samples under consideration.

Given the wide range of concentrations present within the design, limit of detection (LOD) is a practical figure of merit. Limit of detection can be computed as[184]

$$LOD = \frac{k_D \sigma}{m} \tag{5.12}$$

where k_D is the statistical confidence factor (here, $k_D = 3$), m is the slope of a univariate classical least-squares fit of the predicted and reference data, and σ is defined as δr in equation (5.11). Since the predicted versus measured plot was not significantly different from unity, a value of 1.0 was assumed for m in all cases.

5.2.5 Precision Statistics

Precision figures of repeatability and intermediate precision were determined in accordance with ICH guidelines[182], and were reported as the standard deviation of the predicted concentration values (equation 5.8) for repeat measurements. Repeatability and intermediate precision values were established using the randomly-chosen design point comprised of 20% theophylline, 20% lactose, 0% MCC, and 60% starch, compacted at a force of 167.6 MPa. Repeatability, a measure of short-term sampling error, was determined without repositioning of the tablet between successive scans, as well as by removing and subsequently re-centering the compact before acquiring the next measurement. Six scans for each type of repeatability test were collected one after the other on the same day. Intermediate precision, which should incorporate typical variations such as between analysts and days, was determined by scanning the tablet once

a day for six consecutive days. The same analyst collected these scans and the repeatability scans to determine day to day variability of the instruments.

5.3 Results and Discussion

5.3.1 Near-Infrared Analysis

The RMSEC and RMSECV values were plotted against the number of latent variables selected for PLS modeling (Figure 5.2). Savitsky-Golay first derivative preprocessing[185] (11 point smoothing and 2nd order polynomial fit) was chosen based on the minimization of RMSECV. It was independently determined that to adequately model the NIR absorbance data, theophylline required three latent variables, while four were required for lactose, MCC, and starch. Model rank was chosen as the point where a rapid decline in the incremental variance captured was observed, cognizant of the expected feasible limit of dimensionality based on the factors varying within the design. Without derivative preprocessing, an additional latent variable would have been required to compensate for the variation in compact density. It is speculated that the derivative preprocessing most effectively suppressed the physical effect of compaction, which has been shown to have a significant effect on the spectral baseline.[124,161]

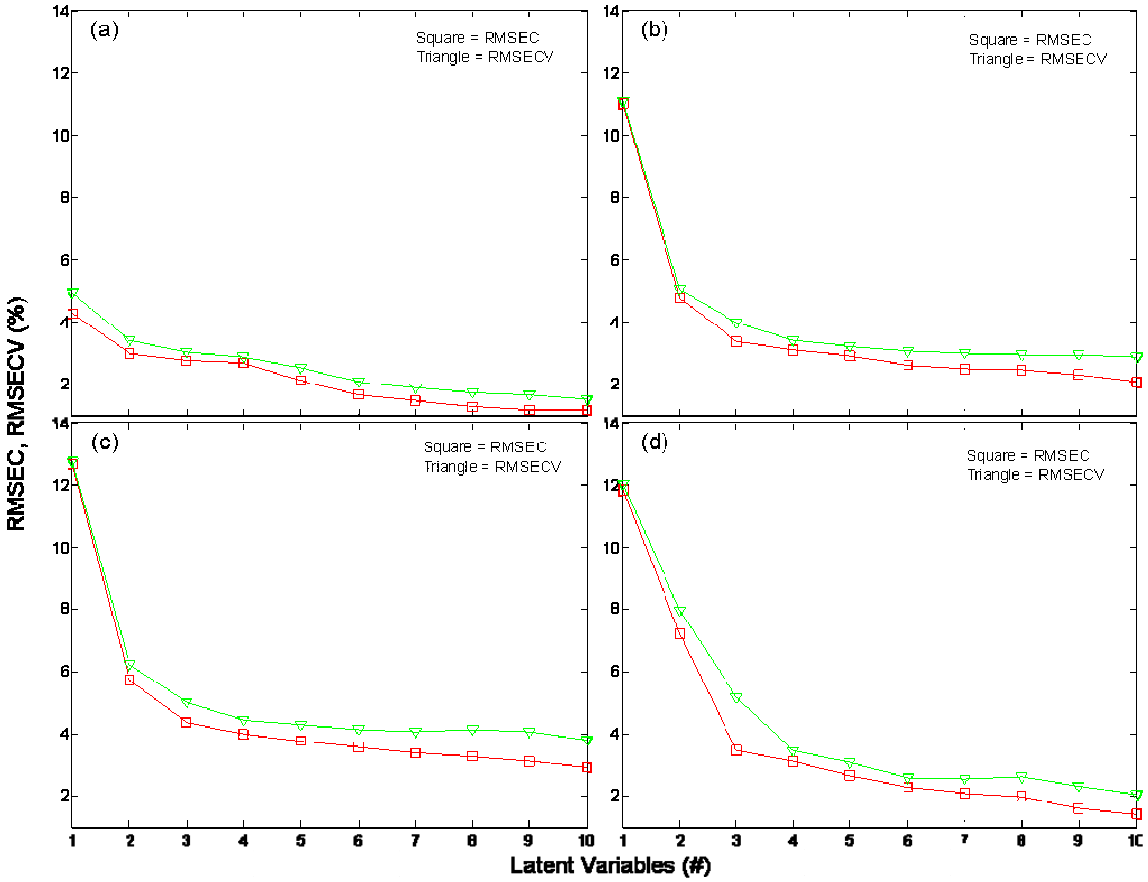

Figure 5.2. Plot of RMSEC (squares) and RMSECV (triangles) versus the number of PLS factors used to model NIR data for theophylline (a), lactose (b), MCC (c), and starch (d), respectively.

Figure 5.3 displays the regression vectors in addition to the pure component spectra for theophylline, lactose, MCC, and starch; note that the pure component spectra and the PLS regression vectors were scaled to facilitate viewing. To gather the pure component scans, powder for each constituent was placed in a glass scintillation vial and spectra were acquired directly through the bottom of the glass; each pure component spectrum represents a mean of three scans. As expected, each regression vector is highly correlated with its associated pure component scan. The goodness-of-fit seen in the predicted versus measured concentration values for the four constituents demonstrates the linearity of the PLS models implemented (Figure 5.4a).

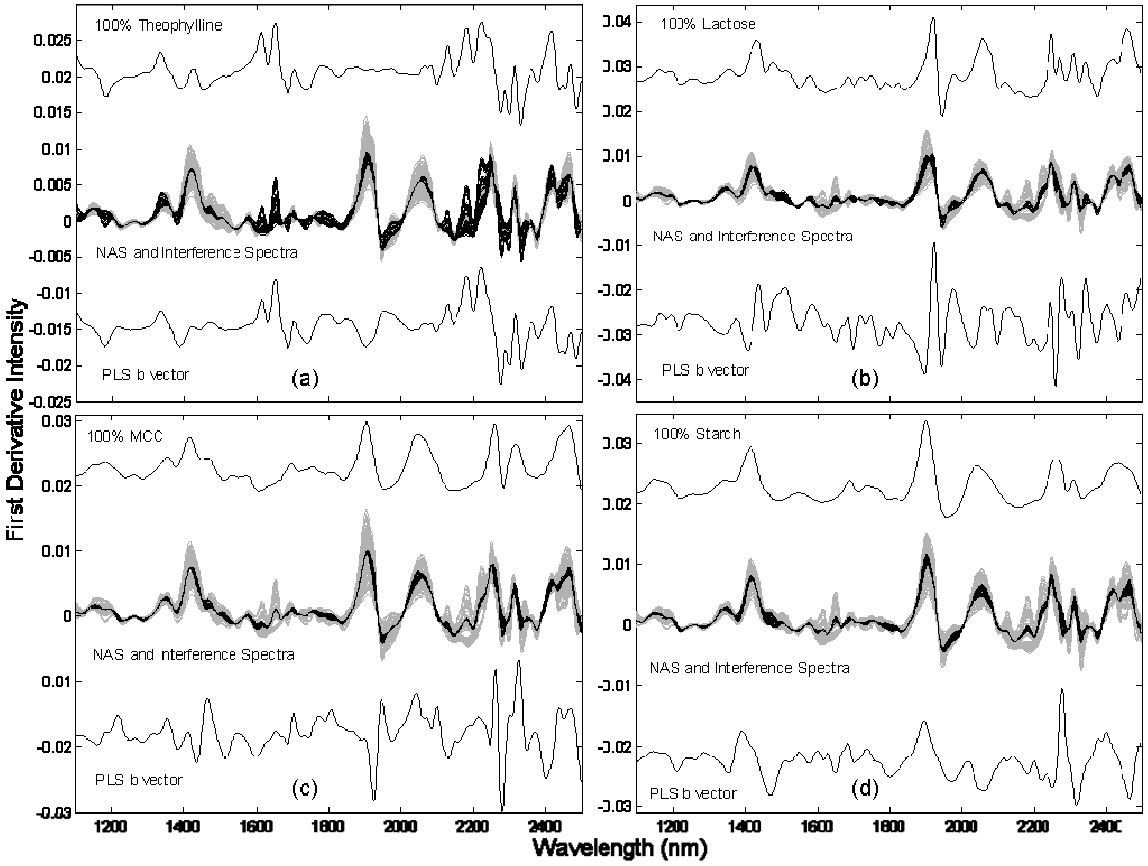

Figure 5.3. NIR pure component spectra (upper solid lines), PLS regression vectors (lower solid lines), net analyte signal (black), and interference (grey) vectors for each calibration sample, for theophylline (a), lactose (b), MCC (c), and starch (d), respectively. Note that the pure component spectra and the PLS regression vectors were scaled to facilitate viewing.

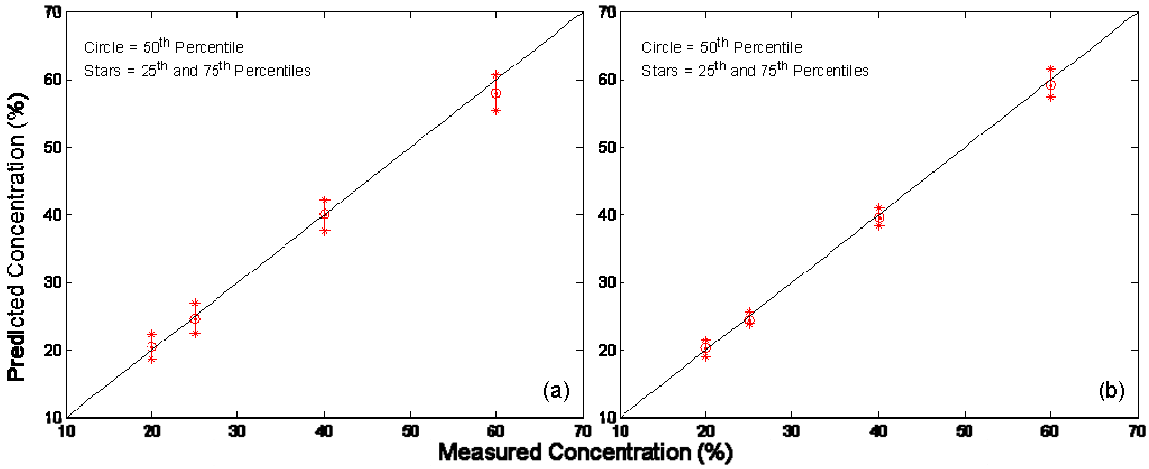

Figure 5.4. Predicted versus measured concentration plot for NIR (a) and Raman (b) data. Circles represent the 50th percentile while the upper and lower stars represent the 25th and 75th percentiles, respectively. The unity line is shown in black.

Following estimation of the PLS regression vectors for each constituent, the portion of the NIR signal related only to the component being analyzed was determined for all calibration samples (equation 5.2). Figure 5.3 exhibits the NAS vector along with the corresponding interference spectrum for each calibration sample for theophylline, lactose, MCC, and starch. This unique plotting scheme directly illustrates the contrast between analyte and interference signals. Furthermore, this graphically illustrates the ability of multivariate calibration to achieve selectivity. For example, in Figure 5.3a, a great deal of spectral variance can be observed around 1500 nm; however, this variation is not attributed to the presence of theophylline, rather, it is the result of the interfering components. This effect is evident by the relatively small range in intensity for the NAS spectra in comparison to the larger intensity range for the interference spectra. Conversely, much of the spectral variation around 1650 nm is due to the variance in theophylline concentration. Similar phenomena can be seen for the other three components (Figure 5.3).

5.3.2 Raman Analysis

The RMSEC and RMSECV were plotted against the number of PLS latent variables modeled in (Figure 5.5). Savitsky-Golay first derivative preprocessing[185] (33 point smoothing and 2nd order polynomial fit), was also selected based on minimization of RMSECV. Presently, there are no known published reports identifying any consistent correlation between variation in Raman spectra and tablet hardness; implying that Raman spectra are insensitive to compact hardness variation. The data collected in this study are in agreement with this conclusion; no discernible pattern was observed relating Raman intensity and tablet compaction force (Figure 5.6). Hence, the role of derivative

preprocessing was apparently not to mitigate any spectral effect of hardness variation; rather, it suppressed the baseline effect present in the spectra. Theophylline and lactose each required four latent variables, while three were required for MCC and starch.

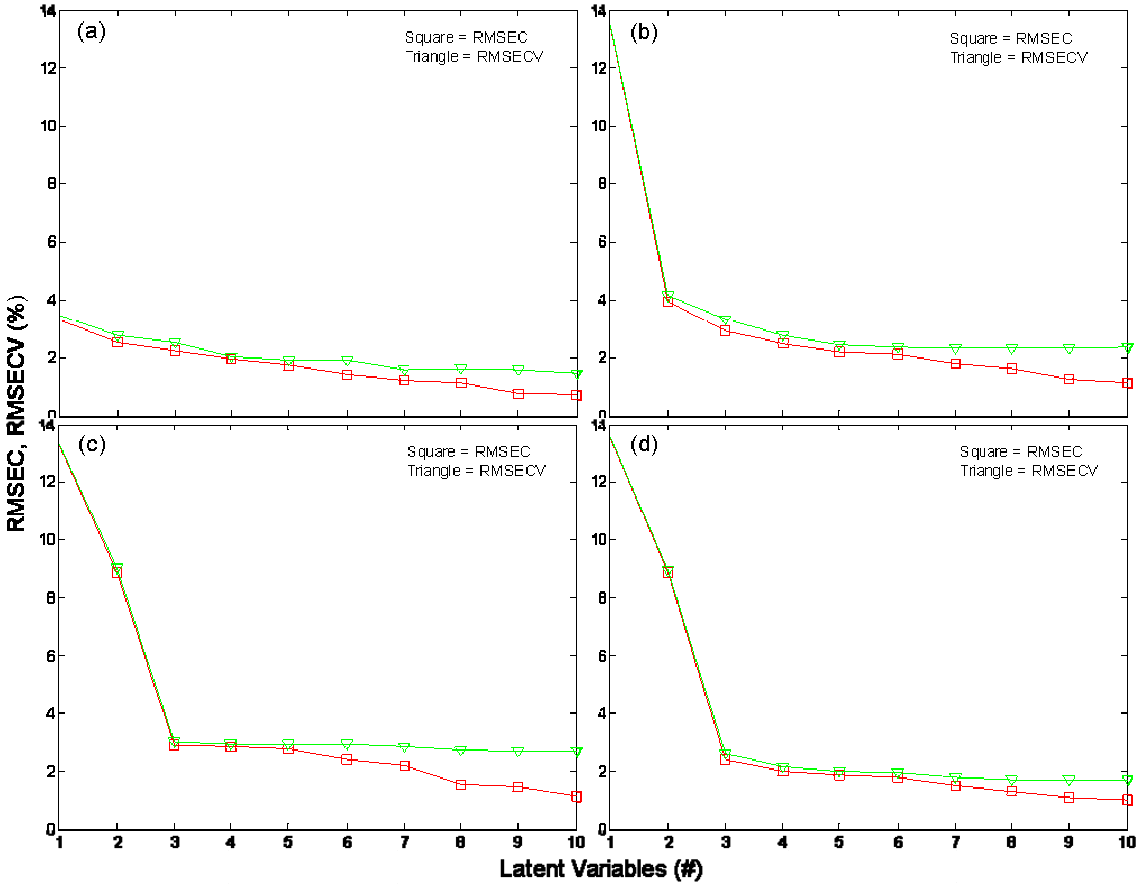

Figure 5.5. Plot of RMSEC (squares) and RMSECV (triangles) versus the number of PLS factors used to model Raman data for theophylline (a), lactose (b), MCC (c), and starch (d), respectively.

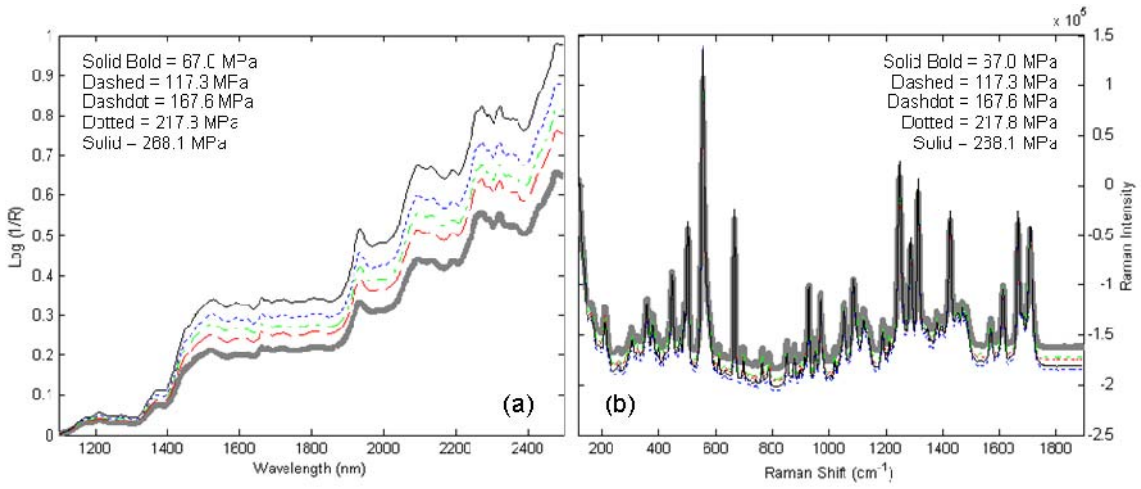

Figure 5.6. NIR (a) and Raman (b) spectra of the same design point (40% theophylline, 40% lactose, 20% MCC, and 0% starch) compacted at 67.0, 117.3, 167.6, 217.8, and 268.1 MPa. For each spectrum, the value for the first variable was subtracted to facilitate viewing.

Figure 5.7 displays the scaled regression vectors in addition to the scaled pure component spectra for theophylline, lactose, MCC, and starch. Raman pure component scans were gathered in the same manner as the NIR. Again, it was anticipated that the PLS regression vectors would include information pertaining to the component, which was confirmed by the similarities between the pure component scans and the regression vectors for all four constituents. Less dispersion in predicted values was observed around each concentration level, which is in agreement with the higher calibration R^2 statistics for the Raman calibration (Figure 5.4).

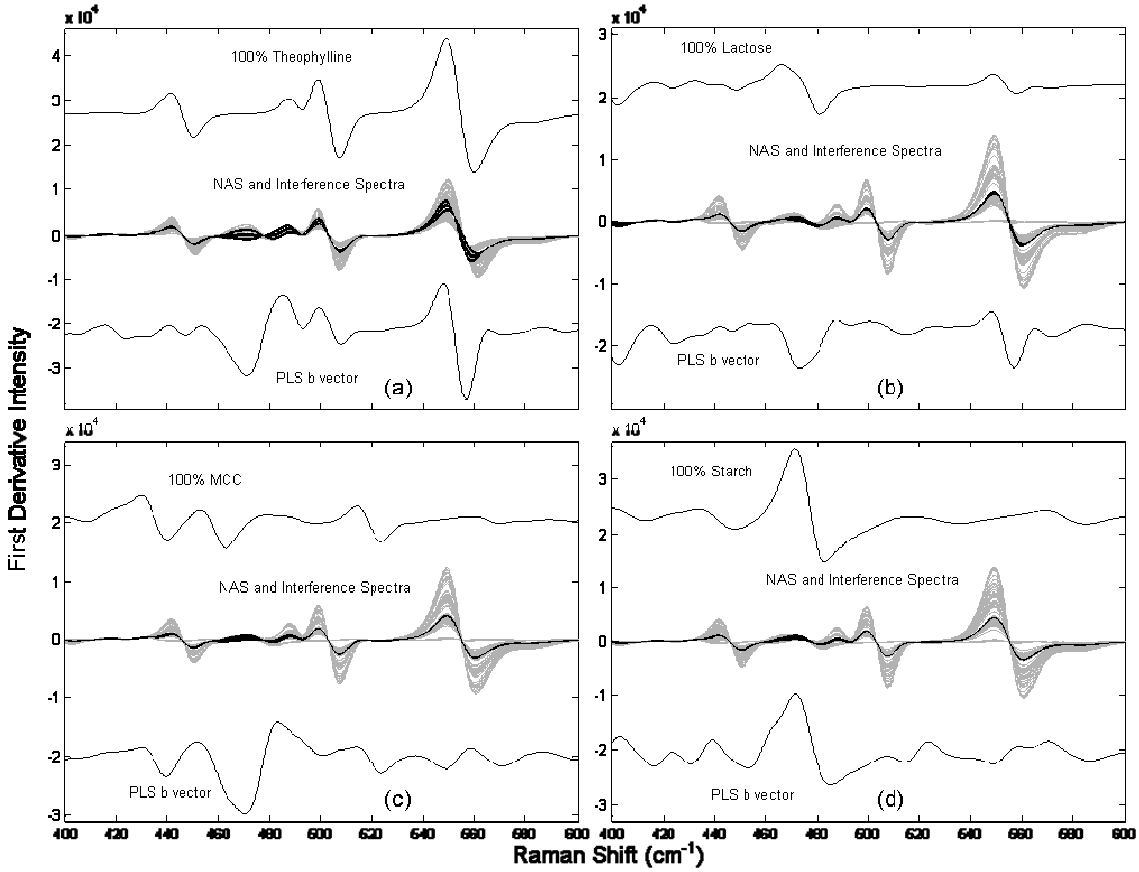

Figure 5.7. Raman pure component spectra (upper solid lines), PLS regression vectors (lower solid lines), net analyte signal (black), and interference (grey) vectors for each calibration sample, for theophylline (a), lactose (b), MCC (c), and starch (d), respectively. Note that the pure component spectra and the PLS regression vectors were scaled to facilitate viewing.

Following the construction of the PLS regression vectors for each constituent, the $\hat{\mathbf{NAS}}_i$ was determined for each calibration sample (equation 5.2). Figure 5.7 depicts the NAS vector and the interference spectrum for each calibration sample for theophylline, lactose, MCC, and starch. Although 205.5 - 1895.7 cm^{-1} was used during calibration, a reduced range was plotted to highlight the contrast between NAS and interference spectra. As was observed for NIR, the patterns demonstrate the selectivity of multivariate calibration for each component.

5.3.3 Figure of Merit Comparison

Table 5.1 displays the calibration statistics describing the performance of the NIR and Raman spectrometers under investigation for this particular four-component solid dosage system. With regard to accuracy, the Raman calibration appears to have an advantage relative to NIR; as reflected by the lower RMSE as well as the superior coefficient of determination. Among the four components, theophylline was modeled the most accurately by both NIR and Raman. This is most likely attributable to two factors: higher sensitivity and selectivity relative to the other components, which will be addressed in greater detail in the following paragraphs.

Table 5.1. Calibration statistics and figures of merit for Reflectance NIR and Raman as determined for each constituent.

Data Type		NIR (Reflectance)				Raman			
Method		PLS							
Preprocessing		1st Derivative (11,2,1)				1st Derivative (33,2,1)			
Latent Variables		3	4	4	4	4	4	3	3
Component		Theophylline	Lactose	MCC	Starch	Theophylline	Lactose	MCC	Starch
Accuracy	R^2 - Cal	0.962	0.951	0.919	0.952	0.981	0.969	0.958	0.972
	R^2 - CV	0.962	0.942	0.902	0.941	0.979	0.962	0.955	0.966
	RMSEC (%)	2.7	3.1	4.0	3.1	1.9	2.5	2.9	2.4
	RMSECV (%)	2.8	3.4	4.4	3.4	2.0	2.8	3.0	2.6
Precision	Repeatability - w/o repositioning (%)	0.01	0.16	0.16	0.02	0.28	0.27	0.09	0.04
	Repeatability - w/ repositioning (%)	0.07	0.10	0.36	0.46	0.45	0.46	0.45	0.27
	Intermediate (%)	0.11	0.16	0.52	0.66	0.35	0.66	0.36	0.26
Sensitivity (Instrument Intensity / %)		0.02	0.01	0.01	0.01	12768.61	5124.58	3732.26	3265.58
Analytical Sensitivity (1 / %)		126.36	82.47	37.09	31.18	15.13	17.31	11.85	10.53
Effective Resolution (%)		0.01	0.01	0.03	0.03	0.07	0.06	0.08	0.09
Selectivity (unitless)		0.59	0.33	0.24	0.27	0.37	0.24	0.18	0.16
Signal-to-Noise (unitless)		282.40	189.88	87.23	72.14	34.99	40.69	26.80	23.98
Limit of Detection (%)		0.33	0.50	1.08	1.31	2.70	2.32	3.53	3.94

The ICH guidance on method validation for repeatability[182] provides a protocol for partitioning sources of variance (i.e., instrumental noise, sample positioning error, instrument drift). In particular, it was noted that repositioning inconsistency was a large contributor to total error for both the NIR and Raman calibrations. For the Raman analyzer approximate 6 mm of the surface (roughly 46 percent of the tablet surface) is sampled; as opposed to the off-line NIR instrument that sampled nearly the entire compact face. Further, NIR has a greater depth of sampling relative to Raman spectroscopy. Therefore, sample heterogeneity and/or sample presentation effects have a greater effect on Raman precision. The impact of sampling heterogeneity is reduced as the number of samples analyzed per time period is increased as a result of averaging. For example, if these methods were implemented in an at-line environment, the difference in precision between Raman and NIR is expected to decrease. Additionally, more equitable comparisons between these two analyzers could be made by implementing a modified sample holder capable of precisely positioning 13 mm tablets, by analyzing tablets of a diameter similar to the Raman sampling size, or by averaging multiple locations on either side of the tablet.

In some cases, error statistics were inconsistent with expected trends, as shown in Table 5.1. For example, intermediate precision values calculated using the Raman data were actually lower than repeatability figures for both MCC and starch. This is unexpected as intermediate precision includes the additional factor of day-to-day instrument drift. This may be indicative of an incomplete estimate of the variance associated with sample repositioning.

While the accuracy and precision data provide a feasible means of comparing these two spectrometers, the power of the evaluation can be enhanced by determining additional calibration figures of merit. Sensitivity for both analytical devices was the largest for theophylline; an increase in its concentration resulted in the greatest response in instrumental intensity. The relative magnitude of peaks in un-scaled pure component spectra (not shown) illustrates this effect. It is important to note that the sensitivity values of individual constituents are not comparable between the two instruments.

Analytical sensitivity is used to compare sensitivity across different measurement technologies. This normalized statistic quantifies sensitivity with respect to analytical precision. Although both devices exhibited the greatest sensitivity for theophylline, the NIR device was more sensitive to all four constituents (in terms of analytical sensitivity). Error of repositioning has a direct effect on σ in equation (5.7), which in turn inflated the denominator, thereby reducing the analytical sensitivity of Raman. The constituent ordering for highest to lowest sensitivity is not identical between the two instruments. This emphasizes the importance of pairing the instrument to the analytical task. The effective resolution results reinforce the results reported for sensitivity. Despite the apparent similarity, this statistic should be considered with respect to quality action limits.

Selectivity is important only when adequate sensitivity is available. A lack of selectivity has the effect of suppressing sensitivity. If adequate sensitivity is not available, improvements of selectivity are futile. Theophylline, which exhibited high relative sensitivity, also exhibited superior selectivity, which is directly attributable to its inherent pure component orthogonality. In contrast, the collinearity among lactose,

MCC, and starch (all being carbohydrates) reduced selectivity. For Raman in particular, the sensitivity (equation 5.6) and selectivity are lower for these components. This is manifested in the performance-related figures of merit (LOD, analytical sensitivity, effective resolution, and S/N). An example of the enhanced interference between these three components can be seen in the NIR at approximately 1500 nm, where the NAS signal is quite large for all three (Figure 5.3).

From the results discussed thus far, the resulting S/N ratio analysis should be straightforward. The NIR calibration included more useable signal in relation to obstructive noise. In some cases, researchers assume that the coefficient of determination is directly predictive of S/N ratio.[186] It is for this reason that technology selection criteria are often based upon R^2 and RMSE, since these statistics are frequently generated during calibration, requiring no additional calculations. The results of this work contradict these assumptions. While Raman outperformed NIR in terms of linearity and accuracy, the S/N ratio for NIR measurements was greater. This occurs because error statistics (R^2 or RMSE) are heavily influenced by the experimental design, while S/N ratio is inherent to the method. Further studies are planned in which calibrations will be optimized according to S/N ratio (as opposed to the traditional method of RMSE); the impetus is to address the aforementioned precision issues, while simultaneously enhancing sensitivity, selectivity, and LOD.

It is interesting to note that despite the lower relative precision of the Raman measurements, which deflates several of the figures of merit, no negative effect on the ability of the SIMPLS algorithm to resolve the covariance structure was observed. This is because inverse least-squares regression is less affected by precision than sample

leverage in the estimation of the true solution. This supports the notion that calibration quality is not sufficient to fully describe method performance.

5.4 Conclusions

This study demonstrates that multivariate figures of merit (determined from net analyte signal theory) can be used to compare calibrations constructed from spectroscopic data collected using two analytical instruments detecting different physical phenomena. The observed calibration performance statistics demonstrate that NIR and Raman are both suitable techniques for the quantitative determination of chemical components within this tablet matrix. Beyond error statistics, multivariate figures of merit provide a clearer assessment of the specific factors limiting performance of the methods while providing a means for general characterization. Furthermore, figures of merit, such as effective resolution and limit of detection, provide an additional mechanism for determining the validity and significance of predicted values upon deployment. For all of these reasons, figures of merit should take a more prominent place among chemometric techniques used in pharmaceutical analytical method development.

Chapter 6: Figures of Merit Comparison of Reflectance and Transmittance Near-Infrared Methods for the Prediction of Constituent Concentrations in Pharmaceutical Compacts

The material presented in Chapter 6 was previously published in Short SM, Cogdill RP, Anderson CA 2008. Figures of Merit Comparison of Reflectance and Transmittance Near-Infrared Methods for the Prediction of Constituent Concentrations in Pharmaceutical Compacts. Journal of Pharmaceutical Innovation 3(1):41 - 50. DOI 10.1007/s12247-008-9020-8.

6.1 Introduction

The Process Analytical Technology (PAT) initiative[6] has increased the performance demands upon analytical methods. In a PAT environment, sensors are implemented for control as opposed to inspection. Process analytical sensors require chemometric modeling/calibration[160] to transform instrument signal into useable information. Multivariate calibration provides a platform for data driven analyzer selection.

Near-infrared spectroscopy (NIRS) has emerged as a useful analytical technique for the nondestructive characterization of solid oral dosage forms. Technology selection is an important early process in method development; following technology selection, the choice of the sampling system is critical for method suitability. Near-infrared (NIR) tablet analyzers are constructed to operate in reflectance and/or transmittance modes.

Measurements of accuracy are often used as the sole discriminating factor in selecting spectrometer configuration.[157,187-207] All but two pharmaceutically-oriented

studies[205,207] determined transmittance NIRS to be superior for the analysis of intact tablets. Most suggest that the enhanced accuracy was due to the larger volume of sample interrogated with the transmittance mode.[201-204] Cogdill *et al*.[207] concluded that, despite the enhanced accuracy, the impact of sampling position error was less for the reflectance mode. Mean-squared error and the coefficient of determination assess agreement with a reference method; multivariate figures of merit characterize the capability of the measurement system.

Multivariate figure of merit, based on net analyte signal (NAS) theory,[170,172] were derived from univariate methods (refer to Olivieri *et al*.[173] for additional information). Net analyte signal theory involves the extraction of a particular portion of signal from the remaining data; it separates information relevant to a particular factor (e.g., chemical constituent) from the residual interfering elements. The NAS is useful for quantification because, by definition, it is orthogonal to the remaining factors within the calibration data. Multiple algorithms are available to estimate the NAS. Comprehensive reviews detailing net analyte signal theory and its mathematics can be found elsewhere.[150,169-173,175-181]

For this work, the net analyte signal was calculated using the formula developed by Bro and Anderson[178]

$$\hat{\mathbf{NAS}}_i = \left(\mathbf{x}_i \cdot \mathbf{b}\right) \cdot \left(\mathbf{b}^T \cdot \mathbf{b}\right)^{-1} \cdot \mathbf{b}^T \qquad (6.1)$$

where, $\mathbf{x}_i$ is a sample spectrum from matrix **X** and **b** is a column vector of the regression coefficients for **X**. All equations herein use **X** to represent spectral matrices; bolded characters other than **X**, e.g., **y**, and **NAS**, represent vectors, and italicized characters, e.g., *x* and *y*, represent scalars. Additionally, the notation ||**x**|| signifies the Euclidean

norm of a vector and the superscript *T* indicates the transpose of a vector. Principal components regression (PCR) or partial least-squares (PLS) regression are common techniques used to estimate **b**.[160] The matrix **X** is generally corrected for the mean; therefore, NAS outputs are mean-centered. Solutions can be rescaled to their original range using the vector of means from the centering operation of **X**. Net analyte signal can also be represented as a scalar using the equation[169]

$$\hat{NAS}_i = \left\| \hat{\mathbf{NAS}}_i \right\| \tag{6.2}$$

This operation does not reduce the information contained by the net analyte signal.[169] Once the NAS has been determined, multivariate figures of merit can be estimated.

The objectives of this work were twofold: (1) to assess the utility of net analyte signal theory for the determination of figures of merit and (2) to apply NAS theory for comparing performance of NIR reflectance and transmittance spectroscopy.

6.2 Materials and Methods

6.2.1 Experimental Design

A fully-balanced, quaternary mixture design comprised of anhydrous theophylline (Lot No. 92577, Knoll AG, Ludwigshafen, Germany), Lactose 316 Fast Flo NF Monohydrate (Lot No. 8502113061, Hansen Labs, New Berlin, WI), microcrystalline cellulose (MCC, Avicel PH-200, Lot No. M427C, FMC BioPolymer, Mechanicsburgh, PA), and soluble starch GR (Lot No. 39362, EMD Chemicals Inc., Gibbstown, NJ) was generated. The approximate median particle size of the theophylline, lactose, MCC, and starch (reported by documentation from their respective suppliers), was 90, 100, 180, and 17 microns, respectively. No further analyses or operations were performed on the

materials to establish or modify the particle size distribution. Twenty-nine design points were chosen to cover a wide range in all constituents and to remove any possibility of factor aliasing (Table 6.1). Analysis of the mixture covariance matrix (not shown) demonstrated the design is balanced in all directions, giving equal emphasis to all constituents.

Table 6.1. Composition design.

Design Point	Anhydrous Theophylline (w/w)	Lactose Monohydrate (w/w)	MCC (PH200) (w/w)	Soluble Starch (w/w)
1	0.600	0.200	0.200	0.000
2	0.400	0.400	0.200	0.000
3	0.200	0.600	0.200	0.000
4	0.400	0.200	0.400	0.000
5	0.200	0.400	0.400	0.000
6	0.200	0.200	0.600	0.000
7	0.600	0.200	0.000	0.200
8	0.400	0.400	0.000	0.200
9	0.200	0.600	0.000	0.200
10	0.600	0.000	0.200	0.200
11	0.400	0.200	0.200	0.200
12	0.200	0.400	0.200	0.200
13	0.000	0.600	0.200	0.200
14	0.400	0.000	0.401	0.200
15	0.200	0.200	0.400	0.200
16	0.000	0.400	0.400	0.200
17	0.200	0.000	0.600	0.200
18	0.000	0.200	0.600	0.200
19	0.400	0.200	0.000	0.400
20	0.200	0.400	0.000	0.400
21	0.400	0.000	0.200	0.400
22	0.200	0.200	0.200	0.399
23	0.000	0.400	0.200	0.400
24	0.200	0.000	0.400	0.400
25	0.000	0.200	0.400	0.400
26	0.200	0.200	0.000	0.600
27	0.200	0.000	0.200	0.600
28	0.000	0.200	0.200	0.600
29	0.250	0.250	0.250	0.250

6.2.2 Compact Production

Materials for each design point mixture were dispensed by weight into 25 mL glass scintillation vials (Data Range, Model No. AX504DR, Mettler Toledo, Columbus, OH) in accordance with the experimental design. The materials were mixed for 5 minute cycles by placing them on the rotating drive assembly of a Jar Mill (US Stoneware, East Palestine, OH, USA). Following each blending iteration, a NIR reflectance spectrum was acquired directly through the bottom of the glass (FOSS NIRSystems 5000, FOSS NIRSystems, Inc., Laurel, MD). Using the constituent concentrations as reference data, an *ad hoc* partial least-squares II (PLS-2) calibration was constructed after each blending cycle to evaluate homogeneity. Mixtures were assumed to be homogeneous when further mixing failed to yield an increase in the calibration's coefficient of determination.

The mixtures from each design point were then subdivided and compacted at 5 pressures (67.0, 117.3, 167.6, 217.8, 268.1 MPa) on a Carver Automatic Tablet Press (Model 3887.1SD0A00, Wabash, IN) using a 13 mm die and flat-faced punches. The dwell time was set to 10 seconds. Six compacts were produced per design point, with the sixth compact's compaction pressure randomly chosen from one of the five possible levels. The compaction order was randomized to minimize heteroscedasticity of experimental error. In total, 174 compacts were produced with a nominal target weight of 800 mg per compact. Preparation and compaction occurred over a three day period, after which the compacts were kept in sealed scintillation vials for 15 days prior to spectroscopic analyses to compensate for any radial and/or axial expansion.

6.2.3 Data Acquisition, Instrumentation, and Software

Near-infrared reflectance measurements for both sides of each compact were acquired using a scanning monochromator instrument, equipped with a Rapid Content Sampler, over the wavelength range of 1100 – 2498 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 5000-II, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD). Prior to spectral acquisition, the compacts were positioned via the centering iris standard on this instrument. Two *ad hoc* partial least-squares II (PLS-2) calibrations, using the constituent concentrations as reference data, were constructed from reflectance spectra corresponding to a particular surface of the compacts. Due to the subtle differences in calibration accuracy, measurements pertaining to only one side of the compacts were used.

Transmittance measurements were acquired on a scanning monochromator instrument equipped with an InSight^{TM} Tablet Analyzer over the wavelength range of 600 – 1898 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 6500, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD). Samples were positioned via a tablet holder suited for compacts ~14.5 mm in diameter. Subsequent to acquisition, the wavelength range of the transmittance spectra was truncated to 800 – 1400 nm.

Reflectance and transmittance data were expressed as log (1/R) and log (1/T), respectively. All spectral data were analyzed in the Matlab environment (version 7.1, The MathWorks, Natick, MA) using the PLS_Toolbox (version 3.0, Eigenvector Research, Inc., Manson, WA) and software routines developed at Duquesne University Center for Pharmaceutical Technology (DCPT).

6.2.4 Partial Least-Squares Analysis

NIR reflectance and transmittance data were analyzed separately but in an identical fashion. Partial least-squares regression[122] was used via the SIMPLS algorithm[123] to relate spectroscopic response to concentration for each constituent on an individual basis. Compacts lacking the particular constituent being analyzed were omitted for the analysis as analyte concentration is incorporated in the denominator of some figures of merit calculations. Thus, the calibration data sets were unique for each analysis.

Preprocessing routines, including standard normal variate (SNV) scaling, detrending, derivatives, and combinations of the preceding were tested.[124] Assessment of the most effective data pretreatment method was made based on a minimization of "batch-wise" cross-validation error, where a batch is defined by a compaction level. Within each cross-validation iteration, compacts produced at a specific pressure were removed, a calibration was constructed, and the concentrations of the omitted samples were predicted by the temporary model. This process was iterated until all of the samples had been predicted, thereby allowing for the determination of the root-mean-standard error (RMSE). The RMSE for cross-validation (RMSECV) and calibration (RMSEC) were calculated using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{6.3}$$

where y_i, $\hat{y}_i$, and n are the measured concentration, the predicted concentration, and the number of samples for the current data set, respectively. According to ICH guidelines,

accuracy expresses the agreeability between reference and predicted values,[208] and thus, RMSEC and RMSECV indicate the accuracy of the particular calibration.

6.2.5 Multivariate Figures of Merit

Upon determining the optimum calibration model, the PLS regression vector can be used to determine the net analyte signal according to equations (6.1) and (6.2). Net analyte signal is the basis for calculation of numerous figures of merit, including sensitivity, analytical sensitivity, selectivity, signal-to-noise ratio, and limit of detection and are estimated for each sample (compact) according to the following formulae.

Sensitivity characterizes signal intensity as a function of analyte concentration. Larger sensitivities signify an enhanced instrumental response to a given change in concentration.[169,170] Sensitivity is calculated as[169,173,177]

$$\hat{\mathbf{SEN}}_i = \frac{\hat{\mathbf{NAS}}_i}{y_i} \tag{6.4}$$

where $\hat{\mathbf{SEN}}_i$ is the vector of sensitivities for each instrument variable, $\hat{\mathbf{NAS}}_i$ is the net analyte signal vector, and y_i is the measured concentration for the i^{th} sample. Sensitivity is reported in units of instrument intensity per concentration. For the work herein, measured concentrations were autoscaled before applying equation (6.4). It is also possible to express sensitivity as a univariate statistic by taking the Euclidean norm of the sensitivity vector

$$\hat{SEN}_i = \| \hat{\mathbf{SEN}}_i \| \tag{6.5}$$

where $\hat{SEN}_i$ is the univariate measure of sensitivity for the i^{th} sample.[177,180] For this work, sensitivity is reported as the mean of the univariate sensitivity values for the particular data set.

Given that equation (6.4) utilizes instrument intensity, sensitivity of multiple devices is directly comparable only when they operate under the same fundamental principles (and units of measure). Considering this limitation, the statistic analytical sensitivity (γ) was developed to provide an impartial assessment between dissimilar analytical techniques.[180] Analytical sensitivity is calculated as

$$\gamma = \frac{\hat{SEN}}{\delta r} \tag{6.6}$$

where $\hat{SEN}$ and δr are the mean of the sensitivity values under consideration found using equation (6.5) and a measure of instrumental noise, respectively. Analytical sensitivity is reported in units of inverse concentration. δr was estimated as the mean standard deviation of $\hat{NAS}_i$ (for the component under consideration) of four compacts widely varying in constituent concentrations. The compacts were scanned once a day for six consecutive days. Additionally, this figure of merit provides a means to estimate the minimum discernible concentration difference for the dynamic range modeled (γ^{-1});[180] this is referred to as effective resolution.

Predicted values were determined according to the equation[183]

$$\hat{\mathbf{y}} = \mathbf{X} \cdot \mathbf{b} \tag{6.7}$$

where **X** represents the spectral matrix and **b** is the PLS regression vector, which is dependent on the number of latent variables applied and the component being considered.

The predicted values were scaled to zero mean and unit variance; consequently, they must be rescaled prior to determining accuracy via equation (6.3). It is important to note that all predicted values are independent of any NAS computations.

Selectivity is a dimensionless univariate measure of the fraction of instrumental signal remaining after accounting for spectral overlap; it is the proportion of signal unaffected by the interfering factors.[169] This metric is calculated for each sample by[169,181]

$$\hat{SEL}_i = \frac{\| \hat{\mathbf{NAS}}_i \|}{\| \mathbf{x}_i \|} \tag{6.8}$$

where $\hat{\mathbf{NAS}}_i$ and $\mathbf{x}_i$ are the NAS vector and the original spectrum for the i^{th} sample, respectively. The magnitude of the selectivity parameter is determined by the degree of spectral interference associated with the particular analyte under consideration. For the work herein, selectivity is reported as the mean of the selectivity values for the current data set.

Signal-to-noise ratio is one of the most important metrics for the general comparison of methods and, using net analyte signal theory, is calculated as[169]

$$\hat{S/N}_i = \frac{\hat{NAS}_i}{\delta r} \tag{6.9}$$

where $\hat{NAS}_i$ is the scalar representation of the NAS vector, and δr was estimated as the mean standard deviation of the predicted concentrations (for the component under consideration) of four tablets broadly varying in constituent concentrations (i.e., the same four tablets previously used). Linear regression was performed between measured concentration and the univariate NAS values to estimate scale (a_1) and offset (a_o)

coefficients to convert NAS scalars to units of concentration. Thus, S/N is a dimensionless metric specific to the current application.

$$\hat{S/N}_i = \frac{a_1 \cdot \left(\hat{NAS}_i \right) + a_o}{\delta r} \tag{6.9a}$$

Signal-to-noise ratio is reported in this document as the mean of the S/N values for all samples under consideration.

Considering the concentration range encompassed by the experimental design, limit of detection (LOD) is a valid figure of merit. Limit of detection can be estimated by[184]

$$LOD = \frac{k_D \sigma}{m} \tag{6.10}$$

where k_D is the statistical confidence factor (here, $k_D = 3$), m is the slope of a univariate classical least-squares fit of the predicted and reference data, and for this work, σ and δr were calculated using the same procedure. A value of 1.0 was assumed for m in all computations given that the slopes of the predicted versus measured plots were not significantly different from unity.

6.2.6 Precision Statistics

Repeatability and intermediate precision were determined according to ICH guidelines.[208] These statistics function as estimates of precision and were reported as the standard deviation of predicted concentrations for the corresponding repeat measurements. Precision statistics were calculated using the randomly-chosen design point comprised of 20% theophylline, 20% lactose, 0% MCC, and 60% starch, compacted at a pressure of 167.6 MPa. Repeatability, a measure of short-term sampling

error, was determined without repositioning of the compact between successive scans, as well as by removing and subsequently re-centering the compact prior to spectral acquisition. Six scans for each repeatability test were collected consecutively on the same day. Intermediate precision, which should incorporate typical variations such as between analysts and days, was assessed by scanning the compact once a day for six consecutive days. All precision data were collected by the same analyst.

6.3 Results and Discussion

6.3.1 Wavelength Selection Criteria

The raw NIR response for all 174 compacts is shown in Figure 6.1 for both the reflectance and transmittance geometries. Mathematical assessment is necessary to establish whether or not the entire wavelength range is useful for the prediction of constituent concentrations. While multiple methods for the selection of optimum wavelength ranges are available (e.g., moving-window algorithms, manual trial and error truncation), this work employed correlation vectors to establish if truncation was warranted. Specifically, the correlation between spectral response and component concentration was assessed for each constituent across all wavelengths; only samples corresponding to the calibration data sets were considered. This procedure yields a full-spectrum correlation vector for each constituent (Figure 6.2).

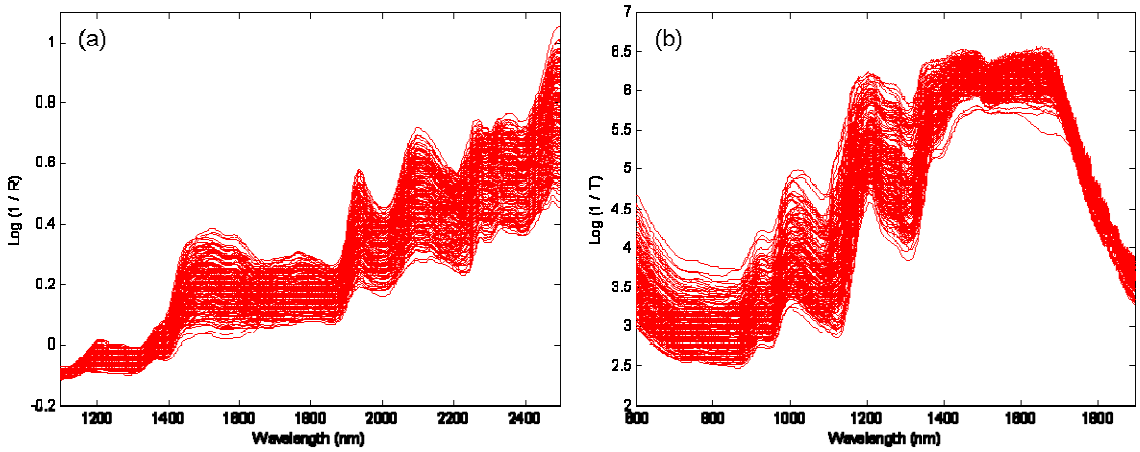

Figure 6.1. Plots of raw reflectance (a) and transmittance (b) spectra for all 174 compacts.

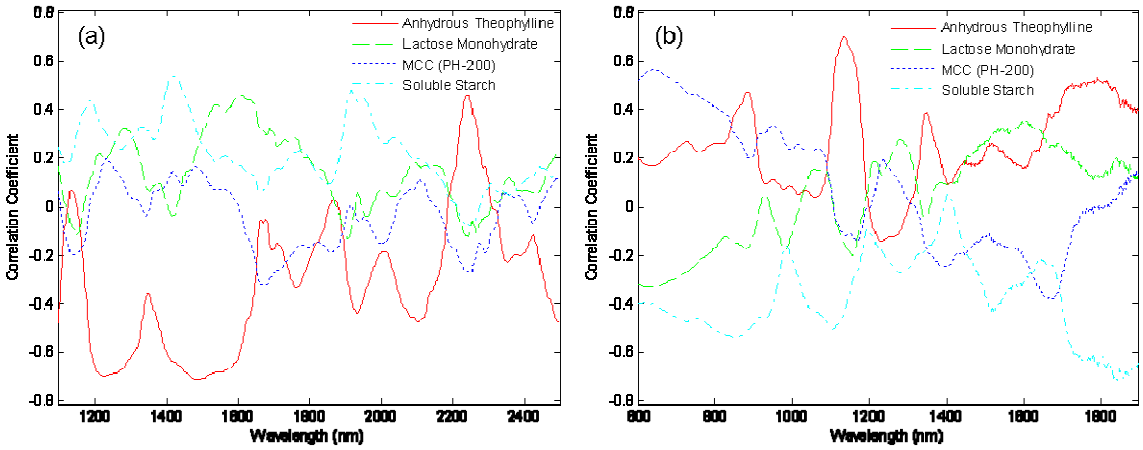

Figure 6.2. Correlation vectors for the reflectance (a) and transmittance (b) instruments illustrating the correlation between instrument intensity and concentration at each wavelength.

Wavelength regions for calibrations should be carefully selected. Figure 6.2 illustrates that the low and high wavelength regions contain meaningful correlations for anhydrous theophylline. Therefore, the entire wavelength range was utilized for the reflectance data. As for the transmittance measurements, Figure 6.2 illustrates that the low and high wavelength extremes are not informative. Furthermore, the transmittance spectra in Figure 6.1 appear to be saturated at higher wavelengths, consistent with the unstable regions of the correlation vector (~ 1400nm). Saturation can be attributed to

sample thickness (mean = 4.387 mm) in combination with the general increase in absorption coefficient that accompanies an increase in wavelength;[209] this results in low transmittance values. For these reasons, the transmittance wavelength range was truncated to 800 – 1400 nm.

6.3.2 Model Development

All predicted values are independent of any net analyte signal mathematics employed. Savitsky-Golay first-derivative preprocessing[185] (eleven point smoothing and second-order polynomial) was used for the reflectance spectra, while a combination of first-derivative preprocessing (three point smoothing and second-order polynomial) and linear detrending was chosen for the transmittance spectra based on the minimization of cross validation error. Model rank was determined (Figure 6.3) using plots of RMSECV and RMSEC versus the number of PLS latent factors. The optimum number of factors was selected based on the inflection point of incremental RMSE (or variance) captured, also considering the agreement between RMSECV and RMSEC values. Model dimensionality was further justified according to the degrees of freedom varying within the experimental design.

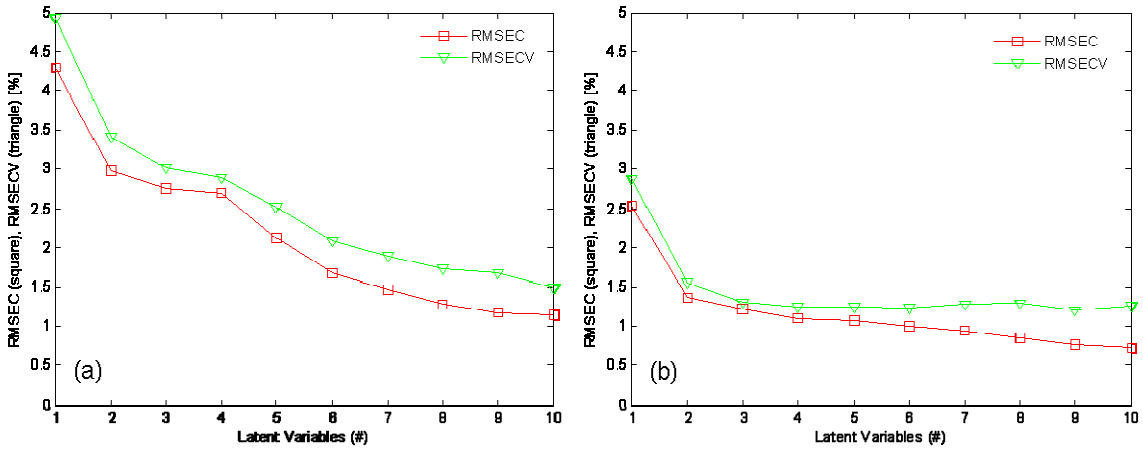

Figure 6.3. Plot of RMSEC and RMSECV versus the number of PLS factors selected to model anhydrous theophylline concentration using reflectance (a) and transmittance (b) spectra.

For both reflectance and transmittance data, three latent variables were required to effectively model anhydrous theophylline concentration, while four latent variables were necessary for lactose monohydrate, MCC, and soluble starch. Calibration is demonstrated by the composite constituent concentration plot (Figure 6.4), which assesses all components simultaneously. The specificity of the calibration models is assessed by comparing the agreement between the regression vectors and the pure component spectra (Figure 6.5). Comparable correlations were observed for the other three constituents (plots not shown).

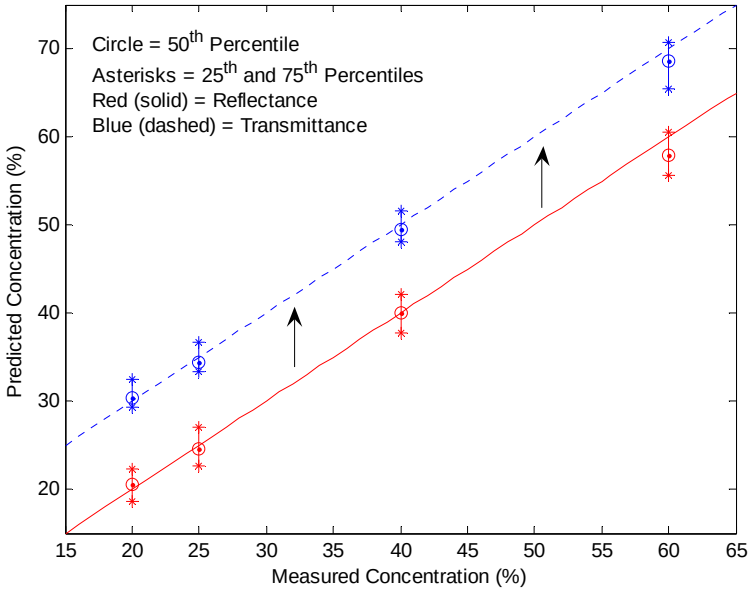

Figure 6.4. Plot of predicted versus measured concentration for reflectance and transmittance data. Circles symbolize the 50th percentile, while the upper and lower asterisks represent the 25th and 75th percentiles, respectively. Transmittance data were offset 10 percent along the ordinate axis to facilitate viewing. The accuracy of all constituents is represented simultaneously.

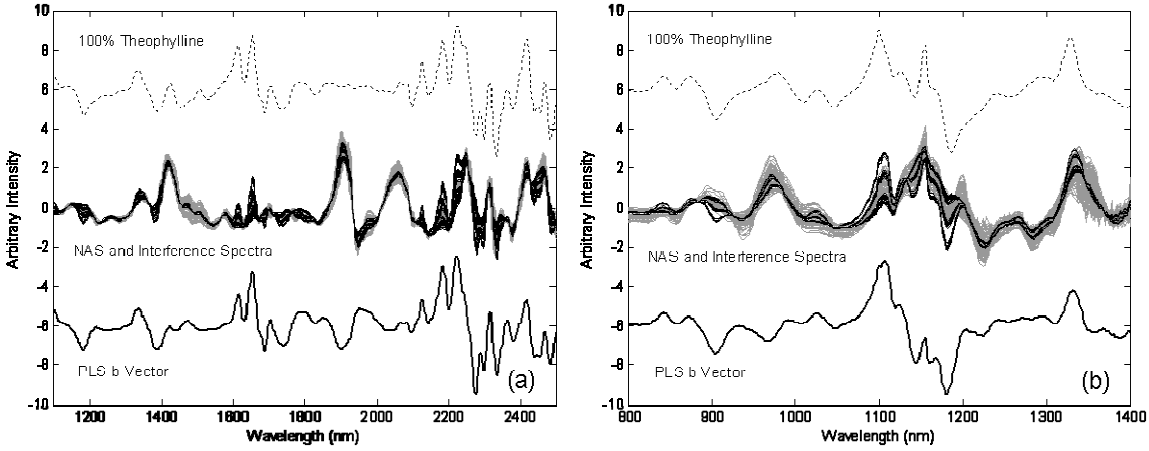

Figure 6.5. Plots containing the pure component spectrum (upper dashed vector), the PLS regression vector (lower bold vector), and the net analyte signal (black) and interference (grey) vectors for each calibration sample, for anhydrous theophylline. Reflectance (a) and transmittance (b) data were scaled to zero mean and unit variance, and were offset, to facilitate viewing.

Subsequent to model selection, the PLS regression vectors were used to determine the net analyte signal for each constituent. The net analyte signal and interference

vectors for each calibration sample are plotted in Figure 6.5 for both the reflectance and transmittance geometries. Contributions of spectral variance, whether from the constituent itself or the other interfering elements, can be observed with this plotting scheme. Additionally, regions enhancing or suppressing the sensitivities, selectivities, and S/N ratios can be examined.

6.3.3 Comparison of Reflectance and Transmittance Methods

The calibration statistics detailing the performance of the reflectance and transmittance geometries for the quantification of constituent concentrations are provided in Table 6.2. Both geometries offer similar results in terms of accuracy. Among the four constituents, the anhydrous theophylline models were most accurate. The most probable explanation for the enhanced accuracy of anhydrous theophylline is its pure component orthogonality, which will be discussed in the following paragraphs. Additionally, the ordering of calibration model accuracy for the different components was identical between the two instruments; however, anhydrous theophylline and lactose monohydrate were more accurately predicted by the transmittance measurements whereas MCC and soluble starch were more accurately assessed by the reflectance measurements. The transmittance method may appear superior, as often times, more importance is placed on the prediction on API concentration, which in this system, is anhydrous theophylline. The negative consequences of this conclusion will be addressed shortly.

Table 6.2. Calibration statistics and figures of merit summarizing method performance for the reflectance and transmittance NIR instruments.

Data Type		Reflectance				Transmittance			
Method		SIMPLS							
Preprocessing		1st Derivative (11,2,1)[a]				1st Derivative (3,2,1)[a] and Detrend			
Latent Variables		3	4	4	4	3	4	4	4
Component		Theophylline	Lactose	MCC	Starch	Theophylline	Lactose	MCC	Starch
Accuracy	R^2 - Cal	0.962	0.951	0.919	0.952	0.992	0.952	0.890	0.928
	R^2 - CV	0.955	0.942	0.902	0.941	0.992	0.933	0.855	0.909
	RMSEC (%)	2.7	3.1	4.0	3.1	1.2	3.1	4.7	3.8
	RMSECV (%)	3.0	3.4	4.4	3.4	1.3	3.7	5.4	4.3
Precision	Repeatability - w/o repositioning (%)	0.01	0.16	0.16	0.02	0.03	0.37	0.34	0.36
	Repeatability - w/ repositioning (%)	0.07	0.10	0.36	0.46	0.11	0.22	0.56	0.80
	Intermediate (%)	0.11	0.16	0.52	0.66	0.12	0.16	0.58	0.46
Sensitivity (Instrument Intensity / %)		0.02	0.01	0.01	0.01	0.10	0.03	0.02	0.03
Analytical Sensitivity (1 / %)		126.36	82.47	37.09	31.18	32.28	39.54	13.34	16.12
Effective Resolution (%)		0.01	0.01	0.03	0.03	0.03	0.03	0.07	0.06
Selectivity (unitless)		0.59	0.33	0.24	0.27	0.61	0.23	0.15	0.20
Signal-to-Noise (unitless)		282.40	189.88	87.23	72.14	74.20	91.32	32.73	36.59
Limit of Detection (%)		0.33	0.50	1.08	1.31	1.27	1.04	2.89	2.58

[a]Paranthetical data corresponds to window width, polynomial order, and derivative order, respectively.

Attempts to partition sources of variance, including instrumental noise, sample repositioning error, and instrumental drift, were made according to the ICH guidance on the validation of analytical procedures.[208] The variation in sample repositioning limits the performance of both modes. Issues of sample inhomogeneity amplify the effects of sample repositioning error; however, its influence can be suppressed through sample averaging. The data in Table 6.2 suggest that the error associated with sample repositioning was not fully captured due to the unanticipated trend in the precision statistics. It is expected that the error of repeatability without repositioning would be lower than repeatability with repositioning, which should be lower than intermediate precision, as each statistic consecutively includes additional factors that are projected to increase precision error. The inclusion of additional data (e.g., scans, repositions, scientists) may help to clarify this issue. Issues such as this must be adequately addressed before comparing methods for possible deployment.

The conventional accuracy and precision data provide a reasonable means of characterizing these two methods; however, this effort can be enhanced with the consideration of multivariate figures of merit. Unlike previous FOM applications where sensitivities were not directly comparable due to dissimilar measurement technologies,[150] the sensitivity values reported in Table 6.2 can be compared, because both the reflectance and transmittance spectrometers measure sample response in absorbance intensity. While the transmittance measurements were more sensitive for all constituents, the order of decreasing sensitivity for both geometries was anhydrous theophylline, lactose monohydrate, soluble starch, and MCC.

Instrumental noise altered the constituent rank ordering for analytical sensitivity. The highest analytical sensitivity for the transmittance method (lactose monohydrate) was roughly comparable to the lowest analytical sensitivity for the reflectance data (soluble starch). This illustrates the power of the analytical sensitivity metric, even when the instruments operate using the same fundamental principles, as the significance of instrumental noise (i.e., sample repositioning) on the transmittance data may have been otherwise overlooked.

The trend for component selectivity mirrors that of sensitivity. It is expected that adequate sensitivity yields acceptable selectivity and equally, inadequate sensitivity generates poor selectivity. In situations where sensitivity is insufficient, attempts to enhance selectivity will be ineffective. The anhydrous theophylline pure component spectrum is the most orthogonal to the interfering (spectral) elements (Figure 6.6). This orthogonality is evident in not only the improved sensitivity, but also the selectivity statistics. The inherent collinearity between lactose monohydrate, microcrystalline cellulose, and soluble starch most likely resulted in the reduced selectivity.

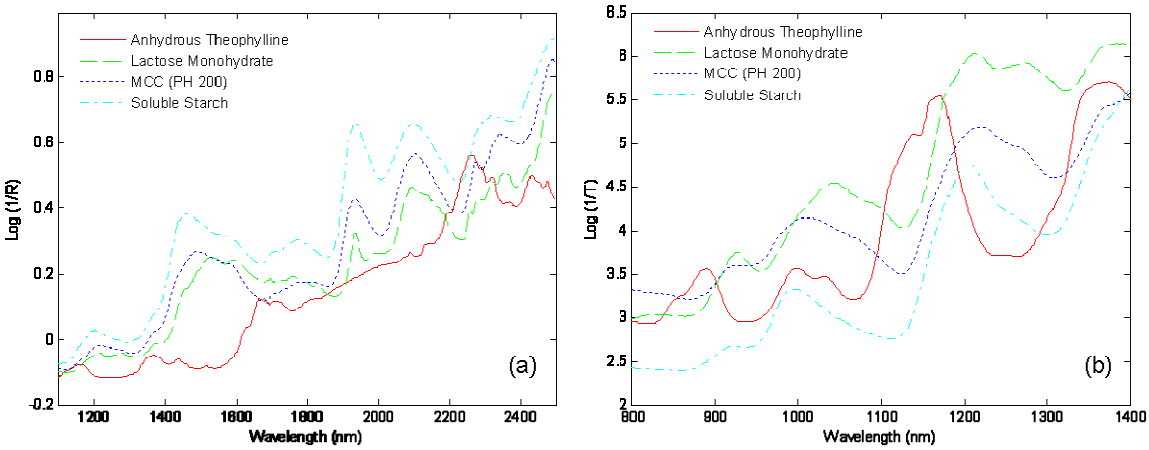

Figure 6.6. Plots of raw reflectance (a) and transmittance (b) pure component spectra. Each spectrum represents the mean response of twenty-seven pure component samples compacted over the range of 67.0 - 268.1 MPa.

The rank order for S/N ratio follows that of analytical sensitivity. Although the transmittance measurements were consistently more selective, the reduction in sample repositioning error (analogous to noise) associated with the reflectance method results in greater S/N ratios for all constituents.

While the statistics summarizing the accuracy of the two methods were generally similar, the disparate performance between the reflectance and transmittance geometries is apparent when considering analytical sensitivity, S/N ratio, and LOD. Upon examination of these statistics, the superiority of the reflectance method for the given application is evident. Although the transmittance method was more accurate and selective for anhydrous theophylline, it was outperformed by the reflectance method in all other figures of merit. Considering only the comparable predictive performance of lactose monohydrate, microcrystalline cellulose, and soluble starch, one may have incorrectly selected the latter method due to its increased accuracy for anhydrous theophylline. The reduction in performance is most likely a consequence of the larger sampling position error observed for the transmittance geometry. It is anticipated that

positioning error could be mitigated by obtaining a tablet holder intended for 13 mm samples. At the time of spectral acquisition, an optimized holder was unavailable. However, it remains a possibility that the transmittance instrument may be inherently noisier (larger δr).

Multivariate figures of merit provide an additional degree of method characterization that cannot be gained from calibration accuracy statistics; however, other factors should be taken into account prior to method deployment. Logistical issues such as cost of implementation, compatibility with existing process/production equipment, and ease of transferability, may dictate the selection. Although a particular method may outperform another in terms of calibration accuracy, ancillary performance and operational issues should be considered when making evaluations.

6.4 Conclusions

The results of this work show that the evaluation of multivariate figures of merit provides a rigorous means of comparing the performance of reflectance and transmittance NIR spectroscopy for the nondestructive prediction of constituent concentrations within compact matrices. The figures of merit analyses revealed performance factors that otherwise may have gone unobserved. Both reflectance and transmittance geometries performed adequately when comparing the calibration accuracy statistics. The precision studies, and more notably the figures of merit, highlight the limitations of the transmittance instrument for this specific system. Performance limitations were most likely attributable to sample positioning error, which was present for both spectrometers, but was more detrimental to the transmittance method. This work also emphasizes the platform that net analyte signal theory provides for the determination of figures of merit,

which are useful for the characterization of calibration performance. While figures of merit are valuable for method characterization, additional logistical factors should be taken into consideration.

Chapter 7: Integration of Process Analytical Technology with Quality by Design to Control the Clinical Performance of a Model Drug Delivery System

7.1 Introduction

The union between PAT and QbD was conceived some years back when the respective initiatives were under development. Although each has an independent role in transitioning from the current to the desired state, the greatest benefits will come from synergistic efforts that integrate PAT and QbD. Together, they are responsible for key tasks in establishing, overseeing, and ensuring pharmaceutical quality.

Recent efforts to remold pharmaceutical quality in terms of risk by relating clinical attributes to production characteristics have transformed approaches to QbD (refer to Chapter 3). While this work presented a comprehensive review of the changes that are to be expected for design space, little discussion was devoted to PAT or its integration within the revised QbD framework. On-, in-, or at-line multivariate sensors are used extensively in a PAT environment, particularly to obtain a greater understanding of the process (and its associated components, including raw and in-process materials), and to ultimately control characteristics of the final product. Applications such as these typically require chemometric modeling (calibration) to transform instrument signal into relevant data (e.g., API/excipient concentration, moisture content, incoming process parameter). Specific information regarding the process (determined via process models) can be used adjust the process (via feedforward or feedback control) to obtain a desired response.

Within the new QbD paradigm, the desired response is final product which demonstrates clinical performance as dictated by the design space. As such, PAT will play an integral role in understanding how (explicitly) the process impacts clinical performance, and, subsequently, how to control the process such that final product meets the safety and efficacy constraints of the design space. The objective of this study was to propose a hypothetical scenario that coupled PAT with QbD such that production could be maintained in a low-risk state.

7.2 Materials and Methods

Data pertaining to 12 distinct wet granulation batches of the model theophylline tablets were utilized to construct the hypothetical scenario. In addition to the 12 and 10 units per batch that were sampled for dissolution and content uniformity testing, respectively, the crushing strength of 10 tablets was estimated for each batch by a diametric crushing test (ElizaTest 3+ Vision Tablet Testing System, Elizabeth-Hata International, North Huntingdon, PA). A reflectance spectrum for both sides of each tablet was acquired before the tablets underwent the relevant destructive analyses. Reflectance was measured over the wavelength range of 1100 – 2498 nm at a 2 nm increment, averaging 32 scans (FOSS NIRSystems 5000-II, Vision version 2.00, FOSS NIRSystems, Inc., Laurel, MD).

Calibration models were constructed using NIR reflectance data, expressed in absorbance units, of the 120 tablets (240 spectra) and the crushing strength measurements (MPa). Partial least-squares regression[122] was used via the SIMPLS algorithm[123] to relate spectroscopic response to crushing strength. Preprocessing routines, including standard normal variate (SNV) scaling, detrending, derivatives, and combinations of the preceding

were tested.[124] The optimal calibration model was selected based on a minimization of cross-validation error. Contiguous block cross-validation with a block size of 5 was used to generate the temporary cross-validation models. Irrespective of the preprocessing routine(s) employed, spectroscopic data were mean-centered while reference data were scaled to zero mean and unit variance. Model rank was chosen as the point where a rapid decline in the incremental variance captured was observed, cognizant of the expected feasible limit of dimensionality based on the factors varying within the design. Calibration efforts revealed that no data pretreatment was necessary. Two latent variables were necessary to adequately model these NIR absorbance data for the prediction of tablet crushing strength. The 288 spectra of the 144 tablets subjected to dissolution testing were projected onto the calibration model to predict their crushing strengths (i.e., testing data set); predictions were compared to the nominal values for each tablet as specified in the original design.[117] Predicted and reference values were used to determine the root-mean-standard error (RMSE). The RMSE for cross-validation (RMSECV), calibration (RMSEC), and testing (RMSET) were calculated using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (7.1)$$

where y_i is the measured crushing strength, $\hat{y}_i$ is the predicted crushing strength, and n is the number of samples for the data set under consideration. The calibration model is summarized in Table 7.1.

Table 7.1. Calibration summary statistics for the prediction of crushing strength.

Data Type	NIR (Reflectance)
Method	PLS
Preprocessing	Raw Spectra
Latent Variables	2
Component	Crushing Strength
R^2_{CAL}	0.836
R^2_{CV}	0.770
R^2_{TEST}	0.926
RMSEC (MPa)	0.30
RMSECV (MPa)	0.37
RMSET (MPa)	0.20
$Bias_{CV}$ (MPa)	0.043
$Bias_{TEST}$ (MPa)	-0.038

These data augment the available information content (i.e., Weibull shape and scale parameters, theophylline content) for the same tablets used in previous studies. As a result of minor spectroscopic differences detected between the two tablet surfaces, the predicted crushing strengths varied slightly for the same tablet. Thus, 288 distinct responses were available for the 144 tablets. These data were then used to determine if an underlying relationship existed between crushing strength and $T_{63.2}$. Modeling efforts utilized standard least-squares regression. Transformations to both the predictor (crushing strength) and the predicted ($T_{63.2}$) variables were explored.

7.3 Results and Discussion

7.3.1 Relationship between Crushing Strength and the Weibull Scale Parameter

Standard least-squares regression revealed a weak correlation between crushing strength and $T_{63.2}$ when all 288 responses were considered simultaneously ($r = -0.3217$). As previous work has demonstrated,[210] however, varying constituent concentrations, as dictated by the experimental design, can mask (spectroscopically and non-

spectroscopically) underlying relationships involving crushing strength (see Chapter 4). When the relationship was re-evaluated with the relative contributions of each component held constant, a much stronger correlation was observed (r = 0.9188); this was determined using 72 of the 288 responses. This was equivalent to assessing one formulation, which was tableted with the distinct purpose of obtaining several target crushing strength values. The function describing the relationship between these two variables was determined to be

$$T_{63.2} = 0.52 \cdot (\textit{Crushing Strength}) - 0.043 \tag{7.2}$$

7.3.2 Integration of PAT with QbD

Simulation revealed that both inefficacy and toxicity risk concerning the model solid oral dosage system were a function of, among other factors, content uniformity and dissolution variability (refer to Chapters 2 and 3). This knowledge was used to generate process-independent design spaces for the model system, which explicitly defined the clinically acceptable combinations of theophylline (% nominal) and $T_{63.2}$ (hours). Based on the findings of the previous studies, the following discussion focuses on integrating PAT at several strategic stages prior to final product release.

Given the relationship between content uniformity and risk, blending of the formulation presents the first viable opportunity to monitor heterogeneity. Numerous non-invasive analytical techniques have proven useful for evaluating blend uniformity (e.g., NIR, Raman). Although uniformity, at least so much as in USP <905>, is solely expressed in terms of the API, process modeling can easily be extended to include various excipients.

Taking advantage of the correlation between crushing strength and $T_{63.2}$, feedforward control can be used to adjust the crushing strength of the final tablets to compensate for the clinical risk associated with various levels of content uniformity. Therefore, the PAT sensor(s) informs the control model (via the process model) of the content uniformity, which then sends a set-point(s) to adjust the compaction pressure (and other pertinent process variables) to modify the release profile and ultimately regulate clinical performance. Since other downstream unit operations have the potential to affect the content uniformity estimated during blending (e.g., wet granulation, milling, tableting) additional calibrations can be deployed to monitor uniformity and/or crushing strength of the product post-tableting. These data can be used to update or even adjust (in real-time) the process and/or control models, as necessary. Various factors, including spectrometer drift, machinery and/or tooling deterioration, or a change in raw materials, could render the use of feedback control invaluable. Ideally, however, the majority of these factors would be accounted for prior to model deployment. The PAT system, therefore, monitors and controls the attributes which are critical-to-quality to ensure that the final product lies within the design space. The integration of PAT and QbD is depicted in Figure 7.1.

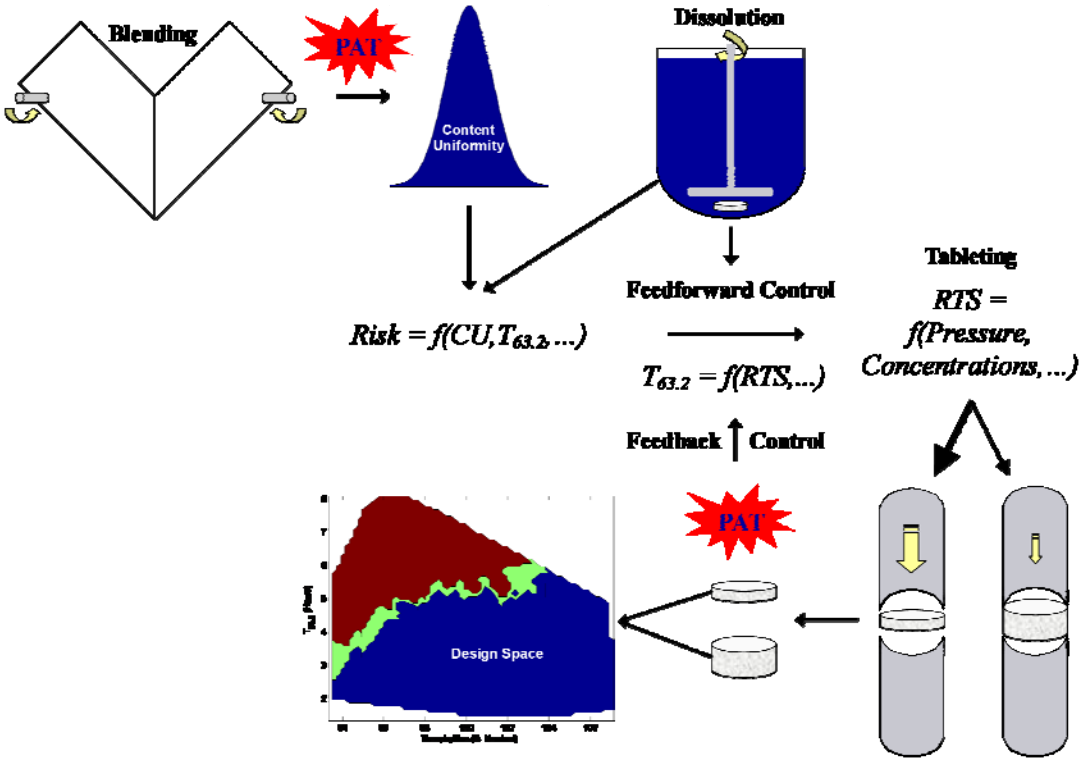

Figure 7.1. Schematic illustrating the integration of PAT with QbD to control clinical performance.

7.4 Conclusions

PAT is an important component of QbD applications. The methodologies developed in the previous phases of this dissertation were used as the basis to propose a hypothetical scenario that coupled PAT with QbD such that production could be maintained in a low-risk state. PAT can be integrated to monitor and control production to ultimately ensure that the critical-to-quality attributes of the final product lay within the design space bounded by clinical risk. Feedfoward control is proposed to adjust tableting such that dissolution would compensate for risk imparted by various levels of content uniformity imposed by upstream processing. Additionally, feedback control could be utilized to retrospectively update or modify the process and/or control models, as needed. Multivariate modeling is an indispensible component of PAT. Future studies

should focus on the effects of model uncertainty and how it propagates through QbD methodologies to ultimately impact estimates of clinical performance.

Chapter 8: Summary

Despite recent advances, the pharmaceutical industry has failed to adequately integrate innovative approaches into its drug discovery and development programs. A 2003 Wall Street Journal article declared that the "pharmaceutical industry has a little secret: Even as it invents futuristic new drugs, its manufacturing techniques lag far behind those of potato-chip and laundry-soap makers." "To the FDA, with its mission of protecting patient safety, it seemed more important to manufacture drugs precisely to specification, using tried-and-true systems, than to latch on to the latest in manufacturing trends." Furthermore, "in other industries, manufacturers constantly fiddle with their production lines to find improvements. But FDA regulations leave drug-manufacturing processes virtually frozen in time."[211] The FDA has since stepped forward to accept partial responsibility for the current state of affairs and introduced or endorsed numerous regulatory initiatives, reports and guidances intended to modernize the Agency's as well as manufacturers' approach to drug discovery and development. This new approach is centered on risk-based decisions.

While the new documents were a colossal step in the right direction, multiple definitions and ambiguous interpretations of "pharmaceutical quality," an element fundamental to every facet of the regulatory, pharmaceutical, and consumer industries, were at odds with the modernization efforts. Mindful of the misnomer, requests were made by a few pioneering individuals to re-define pharmaceutical quality in terms of risk by relating clinical attributes to production characteristics. This work was predicated on this very concept.

The predominant focus of this dissertation was to assemble a platform capable of evaluating the effects of product and patient variability on clinical performance *in silico*. To the best of the author's knowledge, this is the first tool of its kind. Theophylline was selected as the model drug for these studies on the basis of its narrow-therapeutic index and the preponderance of available literature data detailing this compound. Population data specifying asthma rates, anthropometric measurements, and other physiological, pathophysiological and behavioral factors known to influence treatment with theophylline were accrued and used to generate a hypothetical asthmatic patient population. Concurrently, *in vitro-in vivo* correlation, pharmacokinetic, and pharmacodynamic data/models for theophylline were amassed and used to define the architecture of the risk simulation platform.

Specifics regarding the structure of the risk simulation platform were described in great detail. Its framework utilized Monte Carlo simulation and probabilistic risk assessment to evaluate the impact that manufacturing variability had on the clinical performance of model extended-release theophylline tablets. Clinical performance was evaluated through quantitative risk scores for inefficacy and toxicity, which was made possible by utilizing a probabilistic pharmacodynamic model that expressed the probabilities of each adverse event as a function of theophylline plasma concentration.

Production data for a solid oral theophylline dosage system that was formulated, processed, and compacted prior to the conception of this work was used to generate estimates of inter- and intra-batch content uniformity and dissolution variability. These, along with estimates of patient compliance (both percentage of doses consumed and adherence to the dosing regimen), were used to evaluate their impact on clinical

performance in the simulated asthmatic population. A total of 6 factors were studied via two full factorial experimental designs to determine which factor(s) posed a significant risk of inefficacy or toxicity. Sensitivity analyses revealed that patient compliance and intra-batch content uniformity had a significant impact on inefficacy and toxicity risk scores. Intra-batch content uniformity, therefore, was identified as a critical quality attribute. The results of these initial risk simulations demonstrated the conditional risk of manufacturing variability on the clinical performance of the model drug delivery system.

With the framework of the risk simulator solidified and the relationship between product attributes and risk established, the next phase of this research explored the regulatory undertones of re-defining quality. This research initially addressed the potential limitations of the final product specifications utilized in the USP <711> and <905> methods for dissolution and content uniformity, respectively. A series of risk simulations were conducted to highlight the fact that the univariate specifications, upon which the USP methods are centered, disregard potential multivariate and nonlinear interactions that affect clinical performance. The simulations revealed that in-specification product demonstrated an extensive range of inefficacy and toxicity risk scores, magnitudes of which were in excess of those determined to be significant in the sensitivity analyses. Furthermore, the specifications were not consistently sensitive to the lots which offered the utmost safety and efficacy. These results suggested that the final product release tests were unable to adequately appraise final product quality as it related to clinical performance of the model drug delivery system.

The same series of risk simulations were also used to examine the implications that re-casting quality might have on design space initiatives. Contrary to the current

approach to constructing a design space, re-defining quality in terms of risk by relating clinical attributes to production characteristics offers the ability to generate a hyperspace that is explicitly bounded by estimates of clinical performance. Given the new approach to design space, performance-based quality specifications were proposed to be the link between critical quality attributes and clinical performance. The performance-based quality specifications, which are delineated via the risk simulation platform, define a process-independent design space. Design spaces that were conditioned on quantitative estimations of inefficacy and toxicity risk were generated for the model system. Although critical-to-quality attributes are compulsory inputs to a design space, attributes which interact with those determined to be critical should also be included. A design space that defines the relationship between quality attributes and clinical performance can then be applied to specific processes using process models as transfer functions to ultimately link process parameters to clinical performance.

The last segment of this research briefly addressed the role that Process Analytical Technology will play in implementing performance-based quality specifications. Multivariate data analysis has been and will continue to be a key factor in coupling Process Analytical Technology and Quality by Design. Portions of this dissertation explored routines for evaluating the performance of predictive models, particularly calibrations developed using on-, in-, or at-line multivariate sensors. As these types of models undoubtedly will be used to identify, predict, and control elements fundamental to linking manufacturing to clinical performance, their adequacy in terms of predictive performance must be accounted for. Future research should be conducted to investigate

the impact that model uncertainty has on the generation of performance-based quality specifications.

This work described within this dissertation is significant in that it facilitates the transition of the pharmaceutical industry to the desired state. The proposed concepts not only embrace the Process Analytical Technology and Quality by Design initiatives, they support the objectives of the Critical Path Initiative by adopting an innovative approach to pharmaceutical product development. With a direct emphasis on clinical performance, manufacturers and regulators can more easily focus on controlling manufacturing in a manner commensurate with patient risk.

References

1. Cogdill RP, Knight TP, Anderson CA, Drennen III JK 2007. The financial returns of investments in process analytical technology and lean manufacturing: Benchmarks and case study. Journal of Pharmaceutical Innovation 2(1-2):38 - 50.

2. Nunnally BK, McConnell JS. 2007. Six sigma in the pharmaceutical industry: Understanding, reducing, and controlling variation in pharmaceuticals and biologics. 1 ed., Boca Raton: CRC Press. p 204.

3. 1999. Specifications: Test procedures and acceptance criteria for new drug substances and new drug products: Chemical substances Q6A. ICH Harmonised Tripartite Guideline.

4. 2005. Quality risk management Q9. ICH Harmonised Tripartite Guideline.

5. 2004. Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach Final Report. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration.

6. 2004. Guidance for industry: PAT - A framework for innovative pharmaceutical development, manufacturing, and quality assurance. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration.

7. 1978. 21 Code of Federal Regulations Parts 210, Current good manufacturing practice in manufacturing, processing, packaging, or holding of drugs, and 211, Current good manufacturing practice for finished pharmaceuticals. Federal Register: U.S. Department of Health and Human Services, Food and Drug Administration.

8. 2006. Guidance for industry: Quality systems approach to pharmaceutical CGMP regulations. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration.

9. 2004. Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products. Rockville. U.S. Department of Health and Human Services, Food and Drug Administration.

10. 2003. Comparability protocols - Protein drugs products and biological products - Chemistry, manufacturing, and controls information *DRAFT GUIDANCE*. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research, Center for Veterinary Medicine.

11. 2007. Critical path opportunities for generic drugs. U.S. Department of Health and Human Services, Food and Drug Administration.

12. 2008. Pharmaceutical Development Q8(R1). ICH Harmonised Tripartite Guideline.

13. Van Arnum P. 2007. A FDA Perspective on Quality by Design. December 5. PharmTech.

14. 2008. Guidance for industry: Process validation: General principles and practices *DRAFT GUIDANCE*. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research, Center for Veterinary Medicine.

15. 1999. Specifications: Test procedures and acceptance criteria for biotechnological/biological products Q6B. ICH Harmonised Tripartite Guideline.

16. 2006. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, Last Updated: 7/18/2006. In Croarkin C, Tobias P, editors., 2 ed.

17. Nasr MM. 2007. Quality by Design (QbD) - A Modern System Apporach to Pharmaceutical Development and Manufacturing - FDA Perspective. FDA Quality Initiatives Workshop, North Bethesda, Maryland.

18. Williams RL, Adams WP, Poochikian G, Hauck WW 2002. Content uniformity and dose uniformity: Current approaches, statistical analyses, and presentation of an alternative approach, with special reference to oral inhalation and nasal drug products. Pharmaceutical Research 19(4):359 - 366.

19. Woodcock J 2004. The concept of pharmaceutical quality. American Pharmaceutical Review 7(6):10 - 15.

20. Cogdill RP, Drennen JK 2008. Risk-based quality by design (QbD): A Taguchi perspective on the assessment of product quality, and the quantitative linkage of drug product parameters and clinical performance. Journal of Pharmaceutical Innovation 3(1):23 - 29.

21. Dickinson PA, Lee WW, Stott PW, Townsend AI, Smart JP, Ghahramani P, Hammett T, Billett L, Behn S, Gibb RC, Abrahamsson B 2008. Clinical relevance of dissolution tesing in Quality by Design. The AAPS Journal 10(2):280 - 290.

22. 1997. Guidance for industry: Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations. Rockville: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research.

23. Herrmann G, Aynesworth MB 1937. Successful treatment of persistent extreme dyspnea "status asthmaticus": Use of theophylline ethylene diamine (aminophylline, U.S.P.) intravenously. The Journal of Laboratory and Clinical Medicine 23:135 - 148.

24. Heathcote RSA 1920. The action of caffeine, theobromine and theophylline on the mammalian and batrachian heart. The Journal of Pharmacology and Experimental Therapeutics 16(5):327 - 344.

25. Weinberger M, Hendeles L 1996. Drug therapy: Theophylline in Asthma. The New England Journal of Medicine 334(21):1380 - 1388.

26. Moore III LD, Taylor T 1988. Theophylline dosing and theophylline level testing in a family practice population. The Journal of Family Practice 27(1):57 - 61.

27. Barnes PJ 2003. Theophylline: New perspectives for an old drug. American Journal of Respiratory and Critical Care Medicine 167(6):813 - 818.

28. Ginsberg G, Hattis D, Russ A, Sonawane B 2004. Physiologically based pharmacokinetic (PBPK) modeling of caffeine and theophylline in neonates and adults: Implications for assessing children's risks from environmental agents. Journal of Toxicology and Environmental Health, Part A 67(4):297 - 329.

29. Kose N, Yamamoto K, Sai Y, Isawa M, Suwa T, Nakashima E 2005. Prediction of theophylline clearance in CCl_4-treated rats using *in vivo* CYP1A2 and CYP3A2 contents assessed with the PKCYP test. Drug Metabolism and Pharmacokinetics 20(3):168 - 176.

30. Bjorkman S 2004. Prediction of drug disposition in infants and children by means of physiologically based pharmacokinetic (PBPK) modelling: Theophylline and midazolam as model drugs. British Journal of Clinical Pharmacology 59(6):691 - 704.

31. Barnes PJ, Pauwels RA 1994. Theophylline in the management of asthma: Time for reappraisal? European Respiratory Journal 7(3):579 - 591.

32. Weinberger MM 1984. Theophylline QID, TID, BID and now QD? A report on 24-hour dosing with slow-release theophylline formulations with emphasis on analyses of data used to obtain Food and Drug Administration approval for Theo-24. Pharmacotherapy 4(4):181 - 198.

33. Staib AH, Loew D, Harder S, Graul EH, Pfab R 1986. Measurement of theophylline absorption from different regions of the gastro-intestinal tract using a remote controlled durg delivery device. European Journal of Clinical Pharmacology 30(6):691 - 697.

34. Peh KK, Yuen KH 1996. Indirect gastrointestinal transit monitoring and absorption of theophylline. International Journal of Pharmaceutics 139(1-2):95 - 103.

35. Ohnishi A, Kato M, Kojima J, Ushiama H, Yoneko M, Kawai H 2003. Differential pharmacokinetics of Theophylline in elderly patients. Drugs Aging 20(1):71 - 84.

36. Lesko LJ, Tabor KJ, Johnson BF 1981. Theophylline serum protein binding in obstructive airways disease. Clinical Pharmacology and Therapeutics 29(6):776 - 781.

37. Chiou WL, Gadalla MAF, Peng GW 1978. Method for the rapid estimation of the total body drug clearance and adjustment of dosage regimens in patients during a constant-rate intravenous infusion. Journal of Pharmacokinetics and Biopharmaceutics 6(2):135 - 151.

38. Zwillich CW, Sutton FD, Neff TA, Cohn WM, Matthay RA, Weinberger MM 1975. Theophylline-induced seizures in adults: Correlation with serum concentrations. Annals of Internal Medicine 82(6):784 - 787.

39. Hendeles L, Weinberger M 1980. Avoidance of adverse effects during chronic therapy with theophylline. Drug Intelligence and Clinical Pharmacy 14:522 - 530.

40. Sato S, Nakajima M, Honda A, Konishi T, Miyazaki H 2007. Pharmacokinetics of theophylline in guinea pig tears. Drug Metabolism and Pharmacokinetics 22(3):169 - 177.

41. Mitenko PA, Ogilvie RI 1973. Rational intravenous doses of theophylline. The New England Journal of Medicine 289(12):600 - 603.

42. Jacobs MH, Senior RM, Kessler G 1976. Clinical experience with theophylline. Relationships between dosage, serum concentration, and toxicity. The Journal of the American Medical Association 235(18):1983 - 1986.

43. Hendeles L, Bighley L, Richardson RH, Hepler CD, Carmichael J 1977. Frequent toxicity from IV Aminophylline infusions in critically ill patients. Drug Intelligence and Clinical Pharmacy 11:12 - 18.

44. Tateishi T, Asoh M, Yamaguchi A, Yoda T, Okano YJ, Koitabashi Y, Kobayashi S 1999. Developmental changes in urinary elimination of theophylline and its metabolites in pediatric patients. Pediatric Research 45(1):66 - 70.

45. Powell JR, Vozeh S, Hopewell P, Costello J, Sheiner LB, Riegelman S 1978. Theophylline disposition in acutely ill hospitalized patients. The effect of smoking, heart failure, severe airway obstruction, and pneumonia. The American Review of Respiratory Disease 118(2):229 - 238.

46. Jusko WJ, Gardner MJ, Mangione A, Schentag JJ, Koup JR, Vance JW 1979. Factors affecting theophylline clearances: Age, tobacco, marijuana, cirrhosis, congestive

heart failure, obesity, oral contraceptives, benzodiazepines, barbiturates, and ethanol. Journal of Pharmaceutical Sciences 68(11):1358 - 1366.

47. Asmus MJ, Weinberger MM, Milavetz G, Marshik P, Teresi ME, Hendeles L 1997. Apparent decrease in population clearance of theopylline: Implications for dosage. Clinical Pharmacology and Therapeutics 62(5):483 - 489.

48. Abdel-Rahman SM, Kauffman R 2004. The integration of pharmacokinetics and pharmacodynamics: Understanding dose-response. Annual Review of Pharmacology and Toxicology 44:111 - 136.

49. Erdman SM, Rodvold KA, Pryka RD 1991. An updated comparison of drug dosing methods Part II: Theophylline. Clinical Pharmacokinetics 20(4):280 - 292.

50. Sheiner LB, Beal S, Rosenberg B, Marathe VV 1979. Forecasting individual pharmacokinetics. Clinical Pharmacology and Therapeutics 26(3):294 - 305.

51. Yamazaki M, Fukutomi O, Kondo N, Kato Z, Nakashima Y, Shinoda S, Agata H, Kondo T, Imaeda N, Orii T 1994. The design of oral sustained-release theophylline dosing after conversion from intravenous to oral therapy. International Journal of Clinical Pharmacology and Therapeutics 32(11):625 - 631.

52. Sabnis S, Adeyeye CM 1998. Controlled-release hydrophilic tablets for individualized Theophylline therapy. Drug Development and Industrial Pharmacy 25(2):187 - 196.

53. Hurley SF, Linas JD, McNeil JJ, Brignell MJ 1986. A randomized controlled clinical trial of pharmacokinetic theophylline dosing. The American Review of Respiratory Disease 134(6):1219 - 1224.

54. Koup JR, Schentag JJ, Vance JW, Kuritzky PM, Pyszczynski DR, Jusko WJ 1976. System for clinical pharmacokinetic monitoring of theophylline therapy. American Journal of Hospital Pharmacy 33(9):949 - 956.

55. Morgan JN, Sonquist JA 1963. Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association 58(302):415 - 434.

56. Gilman TM, Muir KT, Jung RC, Walberg CB 1985. Estimation of theophylline clearance during intravenous aminophylline infusions. Journal of Pharmaceutical Sciences 74(5):508 - 514.

57. Weinberger M 1987. Theophylline administration. How often is enough? Chest 91(4):636b - 637.

58. Weinberger M 1987. Theophylline dosing frequency: Principles and Practices. The American Review of Respiratory Disease 136(3):793 - 795.

59. Mangura BT, Maniatis T, Abdel Rahman MS, Bartholf R, Lavietes MH 1986. Bioavailability of a once daily-administered theophylline preparation. A comparison study. Chest 90(4):566 - 570.

60. Arkinstall WW, Atkins ME, Harrison D, Stewart JH 1987. Once-daily sustained-release theophylline reduces diurnal variation in spirometry and symptomatology in adult asthmatics. The American Review of Respiratory Disease 135(2):316 - 321.

61. Welsh PW, Reed CE, Conrad E 1986. Timing of once-a-day theophylline dose to match peak blood level with diurnal variation in severity of asthma. The American Journal of Medicine 80(6):1098 - 1102.

62. Bonate PL 2000. Clinical trial simulation in drug development. Pharmaceutical Research 17(3):252 - 256.

63. Rugen P, Callahan B 1996. An overview of Monte Carlo, a fifty year perspective. Human and Ecological Risk Assessment 2(4):671 - 680.

64. Bonate PL 2001. A brief introduction to Monte Carlo simulation. Clinical Pharmacokinetics 40(1):15 - 22.

65. 2001. Risk Assessment Guidance for Superfund: Volume III - Part A, Process for Conducting Probabilistic Risk Assessment. Washington, D.C.: Office of Emergency and Remedial Response, U.S. Environmental Protection Agency.

66. Sobol IM. 1994. A Primer for the Monte Carlo Method. Boca Raton: CRC Press. p 107.

67. Grass GM, Sinko PJ 2002. Physiologically-based pharmacokinetic simulation modelling. Advanced Drug Delivery Reviews 54(3):433 - 451.

68. Holford NHG, Kimko HC, Monteleone JPR, Peck CC 2000. Simulation of clinical trials. Annual Review of Pharmacology and Toxicology 40:209 - 234.

69. Poland B, Wada R 2001. Combining drug-disease and economic modelling to inform drug development decisions. Drug Discovery Today 6(22):1165 - 1170.

70. Gomeni R, D'Angeli C, Bye A 2002. In silico prediction of optimal in vivo delivery properties using convolution-based model and clinical trial simulation. Pharmaceutical Research 19(1):99 - 103.

71. Dutta S, Reed RC 2006. Effect of delayed and/or missed enteric-coated divalproex doses on valproic acid concentrations: Simulation and dose replacement recommendations for the clinician. Journal of Clinical Pharmacy and Therapeutics 31(4):321 - 329.

72. Watanalumlerd P, Christensen JM, Ayres JW 2007. Pharmacokinetic modeling and simulation of gastrointestinal transit effects of plasma concentrations of drugs from mixed immediate-release and enteric-coated pellet formulations. Pharmaceutical Development and Technology 12(2):193 - 202.

73. Wreathall J, Nemeth C 2004. Assessing risk: The role of probabilistic risk assessment (PRA) in patient safety improvement. Quality and Safety in Health Care 13:206 - 212.

74. 2002. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Washington, D.C.: Office of Safety and Mission Assurance, NASA.

75. 1997. Memorandum from Deputy Administrator Fred Hansen on the Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for Monte Carlo Analysis. Washington, D.C.: Office of Research and Development, U.S. Environmental Protection Agency.

76. Kodell RL, Chen JJ, Delongchamp RR, Young JF 2006. Hierarchical models for probabilistic dose-response assessment. Regulatory Toxicology and Pharmacology 45(3):265 - 272.

77. Buur J, Baynes R, Smith G, Riviere J 2006. Use of probabilistic modeling within a physiologically-based pharmacokinetic model to predict Sulfamethazine residue withdrawal times in edible tissues in swine. Antimicrobial Agents and Chemotherapy 50(7):2344 - 2351.

78. Emami J 2006. In vitro - In vivo correlation: From theory to applications. Journal of Pharmacy and Pharmaceutical Sciences 9(2):31 - 51.

79. Uppoor VRS 2001. Regulatory perspectives on in vitro (dissolution)/in vivo (bioavailability) correlations. Journal of Controlled Release 72(1-3):127 - 132.

80. 2009. <711> Dissolution. USP31 NF26. United States Pharmacopeia-National Formulary.

81. 2007. <1088> In vitro and In vivo Evaluations of Dosage Forms. USP30 NF25. United States Pharmacopeia-National Formulary.

82. Shargel L, Yu ABC. 1999. Applied Biopharmaceutics and Pharmacokinetics. 4 ed., Stamford: Appleton & Lange. p 768.

83. Buchwald P 2003. Direct, differential-equation-based in-vitro-in-vivo correlation (IVIVC) method. Journal of PHarmacy and Pharmacology 55(4):495 - 504.

84. Davis SS, Hardy JG, Fara JW 1986. Transit of pharmaceutical dosage forms through the small intestine. Gut 27(8):886 - 892.

85. Agoram B, Woltosz WS, Bolger MB 2001. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Advanced Drug Delivery Reviews 50(Suppl 1):S41 - S67.

86. Wilding I 2000. Site-specific drug delivery in the gastrointestinal tract. Critical Reviews in Therapeutic Drug Carrier Systems 17(6):557 - 620.

87. Yu L, Amidon GL 1999. A compartmental absorption and transit model for estimating oral drug absorption. International Journal of Pharmaceutics 186(2):119 - 125.

88. Amidon GL, Lennernas H, Shah VP, Crison JR 1995. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. Pharmaceutical Research 12(3):413 - 420.

89. Munday DL, Fassihi AR 1995. In vitro-in vivo correlation studies on a novel controlled release theophylline delivery system and on Theo-Dur tablets. International Journal of Pharmaceutics 118(2):251 - 255.

90. El-Yazigi A, Sawchuk RJ 1985. In vitro-in vivo correlation and dissolution studies with oral theophylline dosage forms. Journal of Pharmaceutical Sciences 74(2):161 - 164.

91. Yu Z, Schwartz JB, Sugita ET 1996. Theophylline controlled-release formulations: In vivo-in vitro correlations. Biopharmaceutics and Drug Disposition 17(3):259 - 272.

92. Hussein Z, Friedman M 1990. Release and absorption characteristics of novel theophylline sustained-release formulations: In vitro-in vivo correlation. Pharmaceutical Research 7(11):1167 - 1171.

93. DiPiro JT, Spruill WJ, Blouin RA, Pruemer JM. 2002. Concepts in Clinical Pharmacokinetics. 3 ed., Bethesda: American Society of Health-System Pharmacists, Inc. p 279.

94. Brocks DR, Lee KC, Tam YK, Weppler CP, Bradley JM 1988. A pharmacokinetic dosing method for oral theophylline in pediatric patients. Therapeutic Drug Monitoring 10(1):58 - 63.

95. Casner PR, Reilly R, Ho H 1993. A randomized controlled trial of computerized pharmacokinetic theophylline dosing versus empiric physician dosing. Clinical Pharmacology and Therapeutics 53(6):684 - 690.

96. Weinberger M, Hendeles L, Bighley L 1978. The relation of product formulation to absorption of oral Theophylline. The New England Journal of Medicine 299(16):852 - 857.

97. Takayama K, Morva A, Fujikawa M, Hattori Y, Obata Y, Nagai T 2000. Formula optimization of theophylline controlled-release tablet based on artificial neural networks. Journal of Controlled Release 68(2):175 - 186.

98. Figg WD, McLeod HL. 2004. Handbook of anticancer pharmacokinetics and pharmacodynamics. Totowa: Humana Press, Inc. p 623.

99. Hoffman A, Alfon J, Habib G, Pinto E, Gorodetsky R 1994. The effect of immunosuppression by total-body irradiation on the pharmacodynamics of centrally active drugs in rats. Pharmaceutical Research 11(5):704 - 708.

100. Hoffman A, Pinto E, Afargan M, Schattner A 1994. Cyclosporine enhances theophylline neurotoxicity in rats. Journal of Pharmaceutical Sciences 83(4):559 - 561.

101. Flores-Murrieta F, Mercado DE, Rodriguez-Silverio J, Chapela R 1999. Pharmacokinetic/Pharmacodynamic modeling of theophylline in patients with different degrees of airway obstruction. Proceedings of the Western Pharmacology Society 42:3-4.

102. Thakker KM 1984. Pharmacokinetic-pharmacodynamic modelling and simulation using the electrical circuit simulation program spice2. Biopharmaceutics and Drug Disposition 5(4):315 - 333.

103. Schaefers M, Richter O, Reinhardt D, Becker B 1984. The relationship between pharmacodynamics and pharmacokinetics in asthmatic children receiving a sustained-release formulation of theophylline. International Journal of Clinical Pharmacology, Therapy, and Toxicology 22(8):406 - 415.

104. Andrejak M, Lary H, Aubry P, Levi-Valensi P 1986. Serum levels and bronchodilation after intravenous administration of theophylline in asthmatic patients. Revue de Pneumologie Clinique 42(2):82 - 86.

105. Peters SA 2008. Identification of intestinal loss of a drug through physiologically based pharmacokinetic simulation of plasma concentration-time profiles. Clinical Pharmacokinetics 47(4):245 - 259.

106. Michelson S, Sehgal A, Friedrich C 2006. *In silico* prediction of clinical efficacy. Current Opinion in Biotechnology 17(6):666 - 670.

107. Eddy DM, Schlessinger L 2003. Archimedes: A trial-validated model of diabetes. Diabetes Care 26(11):3093 - 3101.

108. Eddy DM, Schlessinger L 2003. Validation of the Archimedes diabetes model. Diabetes Care 26(11):3102 - 3110.

109. Hallinan Jr. AJ 1993. A review of the Weibull distribution. Journal of Quality Technology 25(2):85 - 93.

110. Justus CG, Hargraves WR, Mikhail A, Graber D 1978. Methods for estimating wind speed frequency distributions. Journal of Applied Meteorology 17:350 - 353.

111. Ahmad AM, Boudinot FD, Barr WH, Reed RC, Garnett WR 2005. The use of Monte Carlo simulations to study the effect of poor compliance on the steady state concentrations of Valproic Acid following administration of enteric-coated and extended release Divalproex Sodium formulations. Biopharmaceutics and Drug Disposition 26(9):417 - 425.

112. 2007. http://www.census.gov/ipc/www/idb/, International Data Base. Washington, D.C.: U.S. Census Bureau.

113. 2007. http://www.cdc.gov/nchs/health_data_for_all_ages.htm, National Center for Health Statistics, Health Data for All Ages. Atlanta: Centers for Disease Control and Prevention.

114. McDowell MA, Fryar CD, Hirsch R, Ogden CL. 2005. Anthropometric Reference Data for Children and Adults: U.S. Population, 1999 - 2002. Advance Data from Vital and Health Statistics, no 361, Atlanta: Centers for Disease Control and Prevention. p 32.

115. 2006. Health, United States, 2006, With Chartbook on Trends in the Health of Americans, no. 2006-1232. Health, United States, Hyattsville: U.S. Department of Health and Human Services, National Center for Health Statistics, Centers for Disease Control and Prevention. p 559.

116. 1996. National Heart, Lung, and Blood Institute Data Fact Sheet: Congestive Heart Failure. Bethesda: U.S. Department of Health and Human Services, National Heart, Lung, and Blood Institute, National Institutes of Health.

117. Voytilla RJ. 2007. Non-destructive prediction of drug release from tablets using near-infrared spectroscopy. Graduate School of Pharmaceutical Sciences, Pittsburgh: Duquesne University. p 212.

118. Umesh B. 1992. Pharmaceutical Dissolution Testing. New York: Marcel Dekker. p 437.

119. Lunney PD, Anderson CA 2009. Investigation of the statistical power of the content uniformity tests using simulation studies. Journal of Pharmaceutical Innovation 4(1):24 - 35.

120. Tsong Y, Shen M, Shah VP 2004. Three-stage sequential dissolution testing rules. Journal of Biopharmaceutical Statistics 14(3):757 - 779.

121. MacGregor JF, Bruwer M-J 2008. A framework for the development of design and control spaces. Journal of Pharmaceutical Innovation 3(1):15 - 22.

122. Geladi P, Kowalski BR 1986. Partial least-squares regression: A tutorial. Analytica Chemica Acta 185:1-17.

123. De Jong S 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometric and Intelligent Laboratory Systems 18:251-263.

124. Cogdill RP, Drennen JK. 2006. Near-Infrared Spectroscopy. In Brittain H, editor Spectroscopy of Pharmaceutical Solids, New York: Taylor & Francis. p 313-412.

125. 2009. U.S. Pharmacopeia & National Formulary. USP31 NF26. Rockville: United States Pharmacopeia-National Formulary.

126. 2009. Theophylline Extended-Release Capsules. USP31 NF26. United States Pharmacopeia-National Formulary.

127. Burns DA, Ciurczak EW editors. 2001. Handbook of Near-Infrared Analysis. 2 ed., New York: Marcel Dekker, Inc. p 814.

128. Kirsch JD, Drennen JK 1995. Determination of film-coated tablet parameters by near-infrared spectroscopy. Journal of Pharmaceutical and Biomedical Analysis 13:1273 - 1281.

129. Drennen JK, Lodder RA. 1993. Pharmaceutical Applications of Near-Infrared Spectrometry. In Patonay G, editor Advances in Near-Infrared Measurements, 1 ed., Greenwich: JAI Press, Inc. p 93 - 112.

130. Kirsch JD, Drennen JK 1999. Nondestructive tablet hardness testing by near-infrared spectroscopy: a new and robust spectral best-fit algorithm. Journal of Pharmaceutical and Biomedical Analysis 19:351 - 362.

131. Morisseau KM, Rhodes CT 1997. Near-infrared spectroscopy as a nondestructive alternative to conventional tablet hardness testing. Pharmaceutical Research 14(1):108 - 111.

132. Blanco M, Alcala M 2006. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy. A contribution to process analytical technologies. Analytica Chimica Acta 557:353 - 359.

133. Tatavarti AS, Fahmy R, Wu H, Hussain AS, Marnane W, Bensley D, Hollenbeck G, Hoag SW 2005. Assessment of NIR spectroscopy for nondestructive analysis of

physical and chemical attributes of sulfamethazine bolus dosage forms. AAPS PharmSciTech 6(1):E91 - E99.

134. Barnes RJ, Dhanoa MS, Lister SJ 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43(5):772 - 777.

135. Isaksson T, Naes T 1988. The effect of multiplicative scatter correction (MSC) and linearity improvements in NIR spectroscopy. Applied Spectroscopy 42:1273 - 1284.

136. Drennen JK. 1990. A "noise" in pharmaceutical analysis: near-infrared outside/inside space evaluation. PhD Thesis: University of Kentucky. p 349.

137. Cogdill RP, Anderson CA, Drennen JK 2004. Using NIR spectroscopy as an integrated PAT tool. Spectroscopy 19(12):104 - 109.

138. Otsuka M, Yamane I 2006. Prediction of tablet hardness based on near infrared spectra of raw mixed powders by chemometrics. Journal of Pharmaceutical Sciences 95(7):1425 - 1433.

139. Blanco M, Alcala M, Gonzalez JM, Torras E 2006. A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity, and dissolution test measurements of intact tablets. Journal of Pharmaceutical Sciences 95(10):2137 - 2144.

140. Chen Y, Thosar SS, Forbess RA, Kemper MS, Rubinovitz RL, Shukla AJ 2001. Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. Drug Development and Industrial Pharmacy 27(7):623 - 631.

141. Donoso M, Kildsig DO, Ghaly ES 2003. Prediction of tablet hardness and porosity using near-infrared diffuse reflectance spectroscopy as a nondestructive method. Pharmaceutical Development and Technology 8(4):357 - 366.

142. Otsuka M 2006. Chemoinformetrical evaluation of granule and tablet properties of pharmaceutical preparations by near-infrared spectroscopy. Chemometric and Intelligent Laboratory Systems 82:109 - 114.

143. Otsuka M, Mouri Y, Matsuda Y 2003. Chemometric evaluation of pharmaceutical properties of antipyrine granules by near-infrared spectroscopy. AAPS PharmSciTech 4(3):Article 47.

144. Guo J-H, Skinner GW, Harcum WW, Malone JP, Weyer LG 1999. Application of near-infrared spectroscopy in the pharmaceutical solid dosage form. Drug Development and Industrial Pharmacy 25(12):1267 - 1270.

145. Kuny T, Schatz C, Ulmschneider M, Marrer S, Leuenberger H. 2003. Non-destructive dissolution testing correlation. Dissolution Technologies, p 22 - 28.

146. Ebube NK, Thosar SS, Roberts RA, Kemper MS, Rubinovitz R, Martin DL, Reier GE, Wheatley TA, Shukla AJ 1999. Application of near-infrared spectroscopy for nondestructive analysis of Avicel powders and tablets. Pharmaceutical Development and Technology 4(1):19 - 26.

147. Gustafsson C, Nystrom C, Lennholm H, Bonferoni MC, Caramella CM 2003. Characteristics of hydroxypropyl methylcellulose influencing compactibility and prediction of particle and tablet properties by infrared spectroscopy. Journal of Pharmaceutical Sciences 92(3):460 - 470.

148. Tanabe H, Otsuka K, Otsuka M 2007. Theoretical analysis of tablet hardness prediction using chemoinformetric near-infrared spectroscopy. Analytical Sciences 23:857 - 862.

149. Shah RB, Tawakkul MA, Khan MA 2007. Process analytical technology. Chemometric analysis of raman and near infra-red spectroscopic data for predicting physical properties of extended release matrix tablets. Journal of Pharmaceutical Sciences 96(5):1356 - 1365.

150. Short SM, Cogdill RP, Anderson CA 2007. Determination of figures of merit for near-infrared and raman spectrometry by net analyte signal analysis for a 4-component solid dosage system. AAPS PharmSciTech 8(4):Article 96.

151. Cogdill RP, Anderson CA, Delgado M, Chisholm R, Bolton R, Herkert T, Afnan AM, Drennen JK 2005. Process analytical technology case study: Part II. Development and validation of quantitative near-infrared calibrations in support of a process analytical technology application for real-time release. AAPS PharmSciTech 6(2):E273 - E283.

152. Otsuka M, Tanabe H, Osaki K, Otsuka K, Ozaki Y 2007. Chemoinformetrical evaluation of dissolution property of indomethacin tablets by near-infrared spectroscopy. Journal of Pharmaceutical Sciences 96(4):788 - 801.

153. Rowe RC, Roberts RJ. 1996. Mechanical Properties. In Alderborn G, Nystrom C, editors. Pharmaceutical Powder Compaction Technology, 1 ed., New York: Marcel Dekker, Inc. p 283 - 322.

154. Zhang Y, Law Y, Chakrabarti S 2003. Physical properties and compact analysis of commonly used direct compression binders. AAPS PharmSciTech 4(4):Article 62.

155. Short SM, Cogdill RP, Anderson CA 2008. Figures of Merit Comparison of Reflectance and Transmittance Near-Infrared Methods for the Prediction of Constituent Concentrations in Pharmaceutical Compacts. Journal of Pharmaceutical Innovation 3(DOI 10.1007/s12247-008-9020-8):41 - 50.

156. Ingle Jr. JD, Crouch SR. 1988. Spectrochemical Analysis. 1 ed., Upper Saddle River: Prentice Hall. p 590.

157. Isaksson T, Miller CE, Naes T 1992. Nondestructive NIR and NIT determination of protein, fat, and water in plastic-wrapped, homogenized meat. Applied Spectroscopy 46(11):1685 - 1694.

158. Dahm DJ, Dahm KD. 2007. Interpreting Diffuse Reflectance and Transmittance. A Theoretical Introduction to Absorption Spectroscopy of Scattering Materials. 1 ed., Chichester, UK: NIR Publications. p 286.

159. Grundy WM, Doute S, Schmitt B 2000. A Monte Carlo ray-tracing model for scattering and polarization by large particles with complex shapes. Journal of Geophysical Research 105(E12):29,291 - 229,314.

160. Martens H, Naes T. 1989. Multivariate Calibration. 1 ed., New York: John Wiley and Sons.

161. Bugay DE, Brittain HG. 2006. Raman Spectroscopy. In Brittain H, editor Spectroscopy of Pharmaceutical Solids, New York: Taylor & Francis. p 271 - 312.

162. Afseth NK, Segtnan VH, Marquardt BJ, Wold JP 2005. Raman and near-infrared spectroscopy for quantification of fat composition in a complex food model system. Applied Spectroscopy 59(11):1324-1332.

163. Furukawa T, Masahiro W, Siesler HW, Ozaki Y 2003. Discrimination of various poly(propylene) copolymers and prediction of their ethylene content by near-infrared and Raman spectroscopy in combination with chemometric methods. Journal of Applied Polymer Science 87:616-625.

164. Nordon A, Meunier C, McGill CA, Littlejohn D 2002. Comparison of calibration methods for the monitoring of a fluorobenzene batch reaction using low-field 19F NMR, 1H NMR, NIR, and Raman spectrometries. Applied Spectroscopy 56(4):515-520.

165. Nordon A, Mills A, Burn RT, Cusick FM, Littlejohn D 2005. Comparison of non-invasive NIR and Raman spectrometries for determination of alcohol content of spirits. Analytica Chemica Acta 548:148-158.

166. Qiao Y, van Kempen TATG 2004. Comparison of Raman, mid, and near infrared spectroscopy for predicting the amino acid content in animal meals. Journal of Animal Science 82:2596-2600.

167. Chung H, Ku M-S 2000. Comparison of near-infrared, infrared, and Raman spectroscopy for the analysis of heavily petroleum products. Applied Spectroscopy 54(2):239 - 245.

168. Ku M-S, Chung H 1999. Comparison of near-infrared and Raman spectroscopy for the determination of chemical and physical properties of naphtha. Applied Spectroscopy 53(5):557-564.

169. Lorber A, Faber K, Kowalski BR 1997. Net analyte signal calculation in multivariate calibration. Analytical Chemistry 69(8):1620-1626.

170. Lorber A 1986. Error propagation and figures of merit for quantification by solving matrix equations. Analytical Chemistry 58:1167-1172.

171. Brown CD 2004. Discordance between net analyte signal theory and practical multivariate calibration. Analytical Chemistry 76:4364-4373.

172. Morgan DR 1977. Spectral absorption pattern detection and estimation. I. Analytical techniques. Applied Spectroscopy 31(5):404-415.

173. Olivieri AC, Faber NM, Ferre J, Boque R, Kalivas JH, Mark H 2006. Uncertainty estimation and figures of merit for multivariate calibration. Pure and Applied Chemistry 78(3):633-661.

174. Haaland DM 1987. Classical Versus Inverse Least Squares Methods in Quantitative Spectral Analyses. Spectroscopy 2:56-57.

175. Boelens HF, Kok WT, de Noord OE, Smilde AK 2004. Performance optimization of spectroscopic process analyzers. Analytical Chemistry 76:2656-2663.

176. Xu L, Schechter I 1997. A calibration method free of optimum factor number selection for automated multivariate analysis. Experimental and theoretical study. Analytical Chemistry 69:3722-3730.

177. Ferre J, Brown SD, Rius FX 2001. Improved calculation of the net analyte signal in inverse multivariate calibration. Journal of Chemometrics 15:537-553.

178. Bro R, Andersen CM 2004. Theory of net analyte signal vectors in inverse regression. Journal of Chemometrics 17(12):646-652.

179. Xu L, Schechter I 1996. Wavelength selection for simultaneous spectroscopic analysis. Experimental and theoretical study. Analytical Chemistry 68:2392-2400.

180. Goicoechea HC, Olivieri AC 2002. Chemometric assisted simultaneous spectrophotometric determination of four-component nasal solutions with a reduced number of calibration samples. Analytica Chemica Acta 453:289-300.

181. Braga JWB, Poppi RJ 2004. Figures of merit for the determination of the polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration. Journal of Pharmaceutical Sciences 93(8):2124-2134.

182. ICH 1997. ICH (International Conference on Harmonisation) Harmonised Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology. Federal Register 62(96):27463-27467.

183. Haaland DM, Thomas EV 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry 60:1193-1202.

184. Long GL, Winefordner JD 1983. Limit of detection: A closer look at the IUPAC definition. Analytical Chemistry 55(7):712A-724A.

185. Savitzky A, Golay MJE 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36(8):1627-1639.

186. Marbach R 2002. On Wiener filtering and the physics behind statistical modeling. Journal of Biomedical Optics 7(1):130 - 147.

187. Ellekjaer MR, Isaksson T 1992. Assessment of maximum cooking temperatures in previously heat treated beef. Part 1: Near infrared spectroscopy. Journal of the Science of Food and Agriculture 59(3):335 - 343.

188. Jeon KJ, Hwang ID, Hahn S, Yoon G 2006. Comparison between transmittance and reflectance measurements in glucose determination using near infrared spectroscopy. Journal of Biomedical Optics 11(1):14022 - 14028.

189. Borjesson T, Stenberg B, Schnurer J 2007. Near-infrared spectroscopy for estimation of ergosterol content in barley: A comparison between reflectance and transmittance techniques. Cereal Chemistry 84(3):231 - 236.

190. Varadi M, Hruschka W, Norris KH 1992. Investigation into the applicability of diffuse reflectance and transmittance technology to tobacco analysis. Acta Alimentaria 21(1):95 - 106.

191. Ellekjaer MR, Hildrum KI, Naes T, Isaksson T 1993. Determination of the sodium chloride content of sausages by near-infrared spectroscopy. Journal of Near Infrared Spectroscopy 1(2):65 - 75.

192. McKenna D 2001. Measuring moisture in cheese by near infrared absorption spectroscopy. Journal of AOAC International 84(2):623 - 628.

193. Thygesen LG 1994. Determination of dry matter content and basic density of Norway spruce by near infrared reflectance and transmittance spectroscopy. Journal of Near Infrared Spectroscopy 2(3):127 - 135.

194. Mitsumoto M, Maeda S, Mitsuhashi T, Ozawa S 1991. Near-infrared spectroscopy determination of physical and chemical characteristics in beef cuts. Journal of Food Science 56(6):1493 - 1496.

195. Segtnan VH, Isaksson T 2000. Evaluating near infrared techniques for quantitative analysis of carbohydrates in fruit juice model systems. Journal of Near Infrared Spectroscopy 8(2):109 - 116.

196. Baye TM, Pearson TC, Settles AM 2006. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. Journal of Cereal Science 43:236 - 243.

197. Kays SE, Shimizu N, Barton II FE, Ohtsubo K 2005. Near-infrared transmission and reflectance spectroscopy for the determination of dietary fiber in barley cultivars. Crop Science 45:2307 - 2311.

198. Scholz E, Prieto-Linde ML, Gergely S, Salgo A, Johansson E 2007. Possibilities of using near infrared reflectance/transmittance spectroscopy for determination of polymeric protein in wheat. Journal of the Science of Food and Agriculture 87:1523 - 1532.

199. Chen H, Marks BP 1997. Evaluating previous thermal treatment of chicken patties by visible/near-infrared spectroscopy. Journal of Food Science 62(4):753 - 756, 780.

200. Rodriguez-Saona LE, Fry FS, McLaughlin MA, Calvey EM 2001. Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. Carbohydrate Research 336:63 - 74.

201. Otsuka M, Tanabe H, Osaki K, Otsuka K, Y. O 2007. Chemoinformetrical evaluation of dissolution property of indomethacin tablets by near-infrared spectroscopy. Journal of Pharmaceutical Sciences 96(4):788 - 801.

202. Schneider RC, Kovar K-A 2003. Analysis of ecstasy tablets: Comparison of reflectance and transmittance near infrared spectroscopy. Forensic Science International 134:187 - 195.

203. Gottfries J, Depui H, Fransson M, Jongeneelen M, Josefson M, Langkilde FW, Witte DT 1996. Vibrational spectrometry for the assessment of active substance in metroprolol tablets: A comparison between transmission and diffuse reflectance near-infrared spectrometry. Journal of Pharmaceutical and Biomedical Analysis 14:1495 - 1503.

204. Thosar SS, Forbess RA, Ebube NK, Chen Y, Rubinovitz RL, Kemper MS, Reier GE, Wheatley TA, Shukla AJ 2001. A comparison of reflectance and transmittance near-infrared spectroscopic techniques in determining drug content in intact tablets. Pharmaceutical Development and Technology 6(1):19 - 29.

205. Merckle P, Kovar K-A 1998. Assay of effervescent tablets by near-infrared spectroscopy in transmittance and reflectance mode: Acetylsalicylic acid in mono and combination formulations. Journal of Pharmaceutical and Biomedical Analysis 17:365 - 374.

206. Williams PC, Sobering DC 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. Journal of Near Infrared Spectroscopy 1(1):25 - 32.

207. Cogdill RP, Anderson CA, Delgado-Lopez M, Molseed D, Chisholm R, Bolton R, Herket T, Afnan AM, Drennen III JK 2005. Process analytical technology case study part I: Feasibility studies for quantitative near-infrared method development. AAPS PharmSciTech 6(2):Article 37.

208. 1997. ICH Harmonised Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology Q2(R1). Federal Register, p 27463-27467.

209. Berntsson O, Burger T, Folestad S, Danielsson L-G, Kuhn J, Fricke J 1999. Effective sample size in diffuse reflectance near-IR spectrometry. Analytical Chemistry 71(3):617 - 623.

210. Short SM, Cogdill RP, Wildfong PLD, Drennen III JK, Anderson CA 2009. A near-infrared spectroscopic investigation of relative density and crushing strength in four-component compacts. Journal of Pharmaceutical Sciences 98(3):1095 - 1109.

211. Abboud L, Hensley S. 2003. Factory shift: New prescription for drug makers: Update the plants; After years of neglect, Industry focuses on manufacturing; FDA acts as a catalyst; The three-story blender. The Wall Street Journal (Eastern edition), September 3, New York. p A.1.