DECISION-SUPPORT ALGORITHMS FOR
BIOPHARMACEUTICAL PORTFOLIO &
CAPACITY MANAGEMENT

A thesis submitted to UCL
for the degree of
ENGINEERING DOCTORATE

by

Edmund Donald George, MEng (Hons)

The Advanced Centre for Biochemical Engineering
Department of Biochemical Engineering
UCL
Torrington Place
London WC1E 7JE

July 2008
I, Edmund George, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
ABSTRACT

Biopharmaceutical drug development is risky, lengthy, and expensive. Decisions in this delicate process are complicated by constraints on resources such as available capacity and uncertainties that include the risk of clinical failure. Hence, the impact of making sub-optimal decisions in this environment can be severe. Accordingly, this work explores the development of algorithms to support strategic drug development decisions and contains four results sections.

Firstly, a decision-support framework based on multi-criteria decision making (MCDM) is presented for assessing options when acquiring biopharmaceutical manufacturing capacity. An example case illustrates the use of this framework where a biopharmaceutical company is faced with options for acquiring commercial manufacturing capacity. The development portfolio consists of three monoclonal antibody drugs at varying stages of clinical development with varying levels of demand. Capacity acquisition options include building in-house capacity, outsourcing, and partnering in addition to some hybrids of these. Deterministic and stochastic analyses showed that building manufacturing capacity ranked highest for the scenario considered when accounting for both financial and operational metrics.

Secondly, the development of a stochastic combinatorial multi-objective optimisation framework is presented which confronts the problem of handling the multitude of decisions and trade-offs when designing portfolio management strategies, which results in extremely large decision spaces. The framework is considerate of strategic decisions that include the portfolio composition, the scheduling of critical development and manufacturing activities, and the involvement of third parties for these activities. The framework simulates development and manufacturing alongside the wider commercial environment. Machine learning and evolutionary computation techniques are also harnessed to characterise the conditional and probabilistic structure of superior decisions and evolve strategies to multi-objective optimality. A case study is constructed to derive insight from the framework where results
demonstrate that a variety of options exist for formulating nondominated strategies in the objective space considered, giving the manufacturer a range of pursuable options. The most preferred means for development across the set of optimised strategies is to fully integrate development and commercial activities in-house, however, alternatives include partnering during early stages of portfolio development and then coordinating outsourced and in-house activities for remaining drugs. Popular scheduling strategies tend to develop two drugs in close succession while spacing out the remaining drug development activities into longer time frames. Thirdly, this framework is expanded to explore the impact of the size of biopharmaceutical drug development portfolio and cash flow constraints on algorithmically formulated strategies. Illustrative examples suggest that naïvely applying strategies optimal for a particular size of portfolio to a portfolio of another size is inappropriate. Also, the size of the portfolio appears to have a larger impact on strategy than the magnitude of cash flow constraint. Fourthly and finally, the economics of biopharmaceutical manufacture are explored with the aim of developing equations that can estimate the cost of manufacturing for both monoclonal antibodies and antibody fragments using mammalian cell culture and bacterial fermentation respectively. The correlations, derived using multiple linear regression, allow the cost of goods to be estimated given the following inputs: the required annual output, fermentation titre, whole process yield, and the probability of achieving a successful batch.
ACKNOWLEDGEMENTS

I would first like to express my gratitude to Dr Suzanne Farid, my academic supervisor, for entrusting me with the independence to research this challenging topic, her assistance in guiding the insightful themes of this work, and her extraordinary eye for detail. I extend my appreciation to Professor Nigel Titchener-Hooker for his original supervisory role in this work and his consistently positive and supportive attitude. I would also like to thank Andrew Sinclair (BioPharm Services, Buckinghamshire, England) for his support.

I gratefully acknowledge the Advanced Centre for Biochemical Engineering (ACBE), support from the Innovative Manufacturing Research Centre for Bioprocessing (IMRC) housed in the ACBE by the Engineering and Physical Sciences Research Council (EPSRC), and financial support in the form of an Engineering Doctorate studentship provided by BioPharm Services and the EPSRC.

I would like to express my sincere appreciation to my friends and colleagues in other settings who have been a great source of learning and encouragement. I especially thank Nadine Grant for her engaging spirit, exemplary successes, and for being one of my greatest counsels for life decisions. I thank Dr Paul Walker and Dr Emma Williams who have provided me with a rare opportunity to enrich personal skillsets that have paid profound dividends both to my work and to the career I have ahead of me. I also express my thanks and admiration to Lorenzo Casarosa, James Maduekeh, Dwayne Philip, and Dr Kafui Tay whose insatiable ambition and entrepreneurial attitudes have been inspirational. Additionally I thank my remaining friends and colleagues at UCL, the Centre for the Advancement of Learning and Teaching, London Business School, the UCL Enterprise Society, Eagles Investors, and the University of London Karate Squad for many constructive and enjoyable experiences.

Finally, I am forever indebted to my family for giving the love, support, and direction that only they can give.
CONTENTS

ABSTRACT 3
ACKNOWLEDGEMENTS 5
CONTENTS 6
LIST OF TABLES 11
LIST OF FIGURES 14

CHAPTER 1. LITERATURE REVIEW 20
1.1 INTRODUCTION 20
1.2. BIOPHARMACEUTICAL DRUG DEVELOPMENT 21
   1.2.1. PROBABILITIES OF SUCCESS 22
   1.2.2. DEVELOPMENT ECONOMICS 25
   1.2.3. DEVELOPMENT TIMELINES 26
1.3. BIOPHARMACEUTICAL MANUFACTURE 26
   1.3.1. CAPITAL COST REQUIREMENTS 31
   1.3.2. COST OF GOODS 34
   1.3.3. DRUG PRICING 36
1.4. OUTSOURCING AND PARTNERING 37
   1.4.1. CONTRACT RESEARCH ORGANISATIONS 38
   1.4.2. CONTRACT MANUFACTURING ORGANISATIONS 40
   1.4.3. PARTNERSHIPS 41
   1.4.4. TIMING THIRD PARTY RELATIONSHIPS 44
1.5. COMPUTATIONAL DECISION MAKING FRAMEWORKS 46
   1.5.1. PORTFOLIO SELECTION AND CAPACITY PLANNING MODELS 46
   1.5.2. FACILITY AND MANUFACTURING PROCESS DECISIONS 49
   1.5.3. STOCHASTIC OPTIMISATION APPROACHES 49
   1.5.4. HANDLING MULTIPLE CRITERIA AND OBJECTIVES 51
1.6. AIMS AND ORGANISATION OF THESIS 52
CHAPTER 2. A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY 55

2.1. INTRODUCTION 55

2.2. DECISION-SUPPORT FRAMEWORK 56
  2.2.1. BIOMANUFACTURING PROCESS MODEL 57
  2.2.2. PROFIT AND LOSS MODEL 60
  2.2.3. CRITERIA 61
  2.2.4. THE WEIGHTED SUM METHOD 64
  2.2.5. UNCERTAINTY 65

2.3. CASE STUDY DESCRIPTION 65
  2.3.1. METHOD 67

2.4. RESULTS AND DISCUSSION 68
  2.4.1. BASE CASE 68
  2.4.1. SENSITIVITY ANALYSIS 70
  2.4.2. EXPECTED NET PRESENT VALUE (ENPV) 70
  2.4.3. STOCHASTIC WSM SCORES 71
  2.4.3.1. RISKS AND REWARDS 72
  2.4.3.2. FINANCIAL AND OPERATIONAL AGGREGATE SCORES 74
  2.4.3.3. OPERATIONAL TO FINANCIAL RATIO 75
  2.4.3.4. SUMMARY OF STOCHASTIC RESULTS 76

2.5. CONCLUSIONS 77

2.6. NOMENCLATURE 78

CHAPTER 3. STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT 81

3.1. INTRODUCTION 81

3.2. MODEL FORMULATION 82
  3.2.1. EVALUATIVE FRAMEWORK 84
  3.2.1.2. STOCHASTIC VARIABLES 88

3.2.2. ESTIMATION OF DISTRIBUTION ALGORITHM 93
  3.2.2.1. FAST NON-DOMINATED SORTING 95
  3.2.2.2. CROWDING DISTANCE 96
  3.2.2.3. K-MEANS CLUSTERING 98
3.2.2.4. Constructing the Bayesian network 99
3.2.2.5. Regenerating the population 102
3.3. Case Study Description 104
3.4. Results 107
  3.4.1. Pareto Front Progression 107
  3.4.2. Analysis of Competing Strategies 111
  3.4.3. Cluster Analysis 115
3.5. Conclusions 121
3.6 Nomenclature 122

CHAPTER 4. STRATEGIC BIOPHARMACEUTICAL PORTFOLIO DEVELOPMENT: AN
ANALYSIS OF CONSTRAINT INDUCED IMPLICATIONS ON STRATEGIES 124
4.1. Introduction 124
4.2. Method 125
4.3. Results and Discussion 126
  4.3.1. Five Drug Portfolio 127
  4.3.1.1. Third Party Strategies 128
  4.3.1.2. Drug Selection Strategies 133
  4.3.1.3. Timing Strategies 134
  4.3.2. Three Drug Portfolio 136
  4.3.2.1. Third Party Strategies 137
  4.3.2.2. Drug Selection Strategies 139
  4.3.2.3. Timing Strategies 140
4.4. Conclusions 141

CHAPTER 5. ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY
MANUFACTURE AND ANTIBODY FRAGMENTS MANUFACTURE 143
5.1. Introduction 143
5.2. Method 145
  5.2.1 Process Overviews 145
  5.2.2 Unit Operations 146
  5.2.3 Cost Outputs 152
  5.2.4 Simulation Method for Data Acquisition 154
  5.2.5 Multiple Regression Technique 155
5.3. RESULTS

5.3.1. FIXED CAPITAL INVESTMENT 157
5.3.2. FERMENTATION AND CHROMATOGRAPHY CAPACITY 161
5.3.3. FIXED INVESTMENT COST PER LITRE 161
5.3.4. COST OF GOODS SOLD PER GRAM 162

5.4. RESULTS: ECONOMIC APPROXIMATIONS 165

5.5. CONCLUSIONS 170

CHAPTER 6. COMMERCIAL CONSIDERATIONS 171

6.1. INTRODUCTION 171

6.2. SOFTWARE ARCHITECTURE 172
6.2.1. GRAPHICAL USER INTERFACE 173
6.2.2. MODELLING ENVIRONMENT 174
6.2.3. SIMULATION PROTOCOLS 175
6.2.4. OPTIMISATION PROTOCOLS 175
6.2.5. DATABASE 176
6.2.6. ANALYTICS 176

6.3. PROJECT IMPLEMENTATION 177
6.3.1. PHASE I: ANALYSIS OF USER REQUIREMENTS 177
6.3.2 PHASE II: SYSTEM DESIGN AND DEVELOPMENT 177
6.3.3. PHASE III: OPERATION AND MAINTENANCE 178
6.3.4 PROJECT COST ANALYSIS 178

CHAPTER 7. VALIDATION 180

7.1. INTRODUCTION 180

7.2. SOFTWARE VERIFICATION AND VALIDATION 181

7.3. REGULATION OF DECISION SUPPORT SYSTEMS 183

CHAPTER 8. CONCLUSIONS AND FUTURE DIRECTIONS 185

8.1. INTRODUCTION 185

8.2. CONTRIBUTIONS OF THIS THESIS 186
8.2.1. A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY 187
8.2.3 Strategic Biopharmaceutical Portfolio Development: An Analysis of Constraint Induced Implications on Strategies 189
8.2.4 Economic Correlations for Whole Monoclonal Antibody Manufacture and Antibody Fragments Manufacture 190
8.2.5. Commercial Considerations for the Development of a Stochastic Combinatorial Multi-Objective Optimisation Framework for Biopharmaceutical Portfolio Management Strategies 190
8.2.6. Validation 191
8.3 Recommendations for Future Work 191
8.3.1. Extending the Probabilistic Model-Building Genetic Algorithm 192
8.3.2. Computing Resources 196
8.3.3. Problem Features 197

REFERENCES 199

APPENDIX A: SUPPORTING DATA FOR CHAPTER 4 216

APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5 218

B.1 INOCULUM 218
B.2 SEED FERMENTATION 219
B.3 Fermentation 221
B.4 Ultrafiltration 224
B.5 Chromatography 226
B.6 Viral Inactivation 231
B.7 Filtration Operations 232
B.8 Centrifugation 235
B.9 Periplasmic Extraction 237
B.10 pH AND CONDUCTIVITY ADJUSTMENT 238
B.11 Dilution 238
B.12 PEGylation 239

APPENDIX C: PAPERS BY THE AUTHOR 240
LIST OF TABLES

CHAPTER 1
Table 1.1. Published clinical trial transition probabilities for biopharmaceutical therapeutics. 24
Table 1.2. Cost data for clinical trials. 25
Table 1.3. Published development time data for biopharmaceuticals (years) developed during 1982-2001. 26
Table 1.4. Construction cost for cGMP biopharmaceutical facilities. 31
Table 1.5. Breakdown of construction costs. 32
Table 1.6. Variables that influence the pharmaceutical cGMP facility construction cost (Pavlotsky, 2004). 33
Table 1.7. Facility cost calculations adapted from Werner (2004). 34
Table 1.8. Annual operating expenses for the production of tPA. 35
Table 1.9. Influence of production rate on specific downstream costs. 36
Table 1.10. Effect of alliances on probability of advancing in a clinical trial. 44

CHAPTER 2
Table 2.1. COGS model breakdown. 59
Table 2.2. Profit and loss model. 60
Table 2.3. Criteria used in the MCDM framework. 63
Table 2.4. Market and development characteristics of drugs in the case study. 66
Table 2.5. Input values used for the case study. 66
Table 2.6. Assumptions for convertible debt financing. 67
Table 2.7. Deterministic results. 68
Table 2.8. Deterministic rankings of the options considered within each decision-making category. 68

CHAPTER 3
Table 3.1. Stochastic variables. 89
Table 3.2. Commercial characteristics of available drug candidates. 105
Table 3.3. Duration and cost information for various phases of the drug development process. 106
Table 3.4. Technical information for manufacture of drug candidates according to clinical phase and corporate body. 106
Table 3.5. Specification of dependencies as related to the number of drugs from the same group within the chosen drug development portfolio. 107
Table 3.6. Specification of contractual dependencies and stage-wise probabilities of success. 107
Table 3.7. Attributes of strategies chosen for comparison. 111
Table 3.8. Structure of the $S1$ portfolio development strategy. 112
Table 3.9. Structure of the $S2$ portfolio development strategy. 113
Table 3.10. Structure of the $S3$ portfolio development strategy. 114
Table 3.11. Structure of the $S4$ portfolio development strategy. 114
Table 3.12. Characteristics of drug selection and timing strategies in each cluster. 115
Table 3.13. Selected probabilities of third party strategies. 119

Chapter 4
Table 4.1. Most probable strategies for a five drug portfolio with no constraints. 130
Table 4.2. Most probable strategies for a five drug portfolio with a constraint of $400MM. 131
Table 4.3. Most probable strategies for a five drug portfolio with a constraint of $250M. 132
Table 4.4. Characteristics of the most probable drug selection and timing strategies for a 5 drug portfolio subject to cash flow constraints. 134
Table 4.5. Most probable strategies for a three drug portfolio with no constraints. 138
Table 4.6. Most probable strategies for a three drug portfolio with a -$100MM constraint. 138
Table 4.7. Most probable strategies for a three drug portfolio with a -$75MM constraint. 139
Table 4.8. Probabilities of drug selection for a three drug portfolio. 140

Chapter 5
Table 5.1. Key inputs and outputs for key operations. 149
Table 5.2. Rules for adding additional equipment units. 151
LIST OF FIGURES

Table 5.3. Lang factor constituents and values. 152
Table 5.4. COGS model breakdown. 153
Table 5.5. Input value ranges. 154
Table 5.6. Additional equipment required for annual demand levels. 160
Table 5.7. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of FCI for whole monoclonal antibody manufacture. 167
Table 5.8. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of FCI for antibody fragment manufacture. 167
Table 5.9. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of COGS/g for whole monoclonal antibody manufacture. 168
Table 5.10. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of COGS/g for antibody fragment manufacture. 168

APPENDIX A

Table A.1. Probabilities of drug selection for a five drug portfolio. 216
Table A.2. Probabilities of drug selection for a three drug portfolio. 217

APPENDIX B

Table B.1. Selected materials and utilities costs. 219
Table B.2. Comparison of predictors for fermenter costs. 224
Table B.3. Relevant equipment price data for the ultrafiltration process. 225
Table B.4. Column volumes, \( v_\alpha \), required for affinity chromatography operations used. 228
Table B.5. Column volumes, \( v_\alpha \), required for anion and cation exchange chromatography operations. 228
Table B.6. Chromatography matrix costs. 229
Table B.7. Chromatography equipment costs. 230
Table B.8. Chromatography buffer costs. 230
Table B.9. Rejection coefficients for components in filtration operations. 233
Table B.10. Relevant equipment price data for the ultrafiltration process. 234
LIST OF FIGURES

CHAPTER 1
Figure 1.1. The drug development process. 21
Figure 1.2. Sales prices of marketed monoclonal antibody drugs. Source: Drug prices adapted from International Pharmaceutical Services (http://www.internationalpharmacy.com), accessed 2004. 37
Figure 1.3. Example of the constituents and possible goal of a portfolio of licensing deals. Source: Adapted from Geyer et al. (1999). 45

CHAPTER 2
Figure 2.1. An overview of the MCDM framework. 56
Figure 2.2. The biomanufacturing process model. Abbreviations: DSP - downstream processing, QC/QA - quality control / quality assurance, FCI - fixed capital investment, COGS - cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. The DSP yield is the overall yield after all downstream processing steps have been completed. The batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values. 58
Figure 2.3. Tornado diagram illustrating the overall maximal effects of best (■) and worst (■) case financial inputs on base case options. 70
Figure 2.4. Distribution of the Expected Net Present Value for the Partner (■), CMO (◇), Build (→), Partner-Build (■ ■) and CMO/Build (☆ ★) options under uncertainty. 71
Figure 2.5. WSM Score of the for the Partner (■), CMO (◇), Build (→), Partner-Build (■ ■) and CMO/Build (☆ ★) options under uncertainty. 72
Figure 2.6. Risk versus reward of considered options. The vertical dotted line and the horizontal dotted line are representative of the mean reward and risk values respectively across all options. The size of each bubble is proportional to the initial expenditure required to exercise each option.

Figure 2.7. Operational versus financial aggregate scores of the Partner (■), CMO (♦), Build (▲), Partner-Build (□) and CMO/Build (○). Each data point is representative of its deterministic value. The x and y error bars each signify one standard deviation either way of the data point. The vertical line intercepting the financial aggregate score axis is the mean financial aggregate score. The horizontal line intercepting the operational aggregate score axis is the mean operational score. The diagonal line shows the profile of financial and operational scores that are equally balanced.

Figure 2.8. The impact of the ratio, R, of operational to financial aggregate weightings on the mean overall aggregate score for the Partner (■), CMO (♦), Build (▲), Partner-Build (□) and CMO/Build (○).

Chapter 3

Figure 3.1. Schematic of the framework used to evaluate populations of candidate solutions and iteratively evolve superior populations of strategies.

Figure 3.2. The super-structure of a candidate strategy for the commercialisation of a portfolio of five drugs.

Figure 3.3. Pseudocode for the EDA.

Figure 3.4. Pseudocode for the nondominated sorting procedure. Adapted from Deb et al. (2002).

Figure 3.5. Pseudocode for the allocation of crowding distances. Source: Adapted from Deb et al. (2002).

Figure 3.6. Pseudocode for the selection of superior candidate solutions.

Figure 3.7. Pseudocode of the k-means clustering algorithm.

Figure 3.8. An example of a Bayesian network with three nodes. δ₁ₓ has no parents and is the parent of δ₂ₓ and δ₃ₓ. δ₂ₓ has one parent, δ₁ₓ, and is the parent of δ₃ₓ. δ₃ₓ has two parents, δ₁ₓ and δ₂ₓ, and no children.
Figure 3.9. Two Bayesian network structures that are disallowed in the model: a. \( \delta_{1,2} \) has a node that is directed to itself; b. \( \delta_{1,2} \) is part of a network structure that is later directed back to itself. 100

Figure 3.10. Pseudocode for Bayesian network construction. 101

Figure 3.11. Pseudocode for generating one candidate solution, \( g \), for the new population \( G(t+1) \) from \( B_z \). 103

Figure 3.12. Progressive discovery of the objective space: a. 1st generation, b. 3rd generation, c. 7th generation, d. 10th generation, e. 14th generation, f. 17th generation. 108

Figure 3.13. Mean positive NPV versus \( p(\text{NPV}>0) \). Note: ( — ) represents an estimate of the non-dominated frontier. Strategies chosen for further analysis are annotated as S1, S2, S3, and S4. 110

Figure 3.14. Pareto fronts for Mean positive NPV versus \( p(\text{NPV}>0) \). Note: ( — ) is the Pareto front discovered by the EDA, and ( — — ) is the Pareto front discovered by random search. 111

CHAPTER 4

Figure 4.1. Example of NPV over time for the development of a portfolio of biopharmaceutical drugs. 126

Figure 4.2. Mean positive NPV versus \( p(\text{NPV}>0) \) for a five-drug portfolio under the following constraint levels: unconstrained ( - ), -$200MM ( ▲ ), -$100MM( ■ ), -$75MM( ● ). 128

Figure 4.3. Mean positive NPV versus \( p(\text{NPV}>0) \) for a three-drug portfolio under the following constraint levels: unconstrained ( - ), -$100MM( ▲ ), -$75MM( ● ). 137

CHAPTER 5

Figure 5.1. The whole monoclonal antibody biopharmaceutical manufacturing process model. Abbreviations: DSP - downstream processing, QC/QA - quality control / quality assurance, FCI - fixed capital investment, COGS - cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. DSP yield is the overall yield after all downstream processing steps have
been completed. Batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.

Figure 5.2. The antibody fragments biopharmaceutical manufacturing process model. Abbreviations: DSP - downstream processing, QC/QA - quality control / quality assurance, FCI - fixed capital investment, COGS - cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. DSP yield is the overall yield after all downstream processing steps have been completed. Batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.

Figure 5.3. Facility size versus cost relationship based on data from Pavlotsky (2004) and Farid (2007). The solid line represents the line of best fit as determined by geometric regression. The equation of this line is $f(x) = 1730x^{0.9136}$, with $R^2 = 0.9539$.

Figure 5.4. Fixed capital investment (£MM) against annual demand for production of: (a) mammalian cell-culture derived monoclonal antibodies, (b) E.coli derived antibody fragments. Titre = 1g/L, yield = 0.6, and $p_B = 0.9$.

Figure 5.5. Fixed USP (—) and DSP (— —) equipment costs (£MM) against annual demand for: (a) mammalian cell-culture derived monoclonal antibodies, (b) E.coli derived antibody fragments. Titre = 1g/L, yield = 0.6, and $p_B = 0.9$. Equipment cost figures include all equipment required for unit operations and their ancillary equipment such as holding tanks.

Figure 5.6. Fermentation capacity (—) and chromatography capacity (— —) requirements against annual demand for: (a) mammalian cell-culture derived monoclonal antibodies, (b) E.coli derived antibody fragments. Fermentation capacity is the total fermentation capacity required across all fermenters. Chromatography
capacity is the volume of all resins used across all chromatography unit operations. Titre = 1 g/L, yield = 0.6, and $p_B = 0.9$.

Figure 5.7. Fixed capital investment per litre of fermentation capacity against fermentation capacity for: (a) mammalian derived whole monoclonal antibodies manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = 1 g/L, yield = 0.6, and $p_B = 0.9$.

Figure 5.8. COGS/g against annual demand for: (a) mammalian derived whole monoclonal antibody manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = 1 g/L, yield = 0.6, and $p_B = 0.9$.

Figure 5.9. The percentage that direct raw materials(---), indirect costs (----), and labour (-----) comprise COGS/g against annual demand for: (a) mammalian derived whole monoclonal antibody manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = 1 g/L, yield = 0.6, and $p_B = 0.9$.

Figure 5.10. Error analysis for FCI approximation for mammalian-derived whole antibody manufacture. Where $T = 1$ g/L and $p_B = 0.9$.

Figure 5.11. Error analysis for COGS/g approximation for mammalian-derived whole antibody manufacture. Where $T = 1$ g/L and $p_B = 0.9$.

Figure 5.12. Error analysis for FCI approximation for bacteria-derived antibody fragments manufacture. Where $T = 1$ g/L and $p_B = 0.9$.

Figure 5.13. Error analysis for COGS/g approximation for bacteria-derived antibody fragments manufacture. Where $T = 1$ g/L and $p_B = 0.9$.

**Chapter 6**

Figure 6.1. Schematic of software architecture.

Figure 6.2. Cost and time breakdown for an example project.

**Appendix B**

Figure B.1. Cost versus volume for a fermenter. A logarithmic regression fit to the data is also shown where $f(x) = 73254\ln(x) - 64605$, and correlation coefficient is 0.9002. Data source: Personal discussion with C.Osborn at the ACBE, UCL (2004), and Jacobs (2006).

Figure B.2. Cost versus volume relationship for holding tanks. The solid line represents the line of best fit as determined by geometric regression. The equation of
this line is \( f(x) = 244.33x^{0.6087} \), with \( R^2 = 0.9988 \). Data source: Personal discussion with C.Osborn at the ACBE, UCL (2004), and Jacobs (2006).
CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

Biopharmaceutical developers face highly significant challenges from the commencement of research and development through to the marketing stage of a therapeutic. The cost of drug development has been reported to be in excess of $800MM (DiMasi et al., 2003; Adams and Bratner, 2006; Boston Consulting Group, 2001); with development times from clinic to market typically being 6 to 10 years (Foo et al., 2001; Ashton, 2001; Werner, 2004; Reichert, 2003); and probabilities of success usually being below 35% (Pavlou and Reichert, 2004; Reichert et al., 2005). Additionally, Ransohoff (2004) comments that biopharmaceutical products are also among the most technically complex to manufacture of any industry. These key points collectively convey that planning for successful biopharmaceutical development is non-trivial and that developers are forced to make optimal use of their resources through intelligent and thorough planning. Despite this Dabbah (2007) indicates that relatively few pharmaceutical and biotechnological organisations have been able to properly manage the interface between research and development and their overall organisation. This suggests that there is a significant opportunity for comprehensive portfolio management tools to assist industrial decision makers towards more effective strategy formulation and execution. Consequently, the underscored aim of this thesis is the development of a drug development portfolio management tool that is capable of simulating and optimising strategic decision making in this challenging setting.

To establish a starting point for the development of decision-support algorithms for biopharmaceutical drug development this introductory chapter explores published literature on the drug development process, drug manufacture, alternative options to in-house development, and computational frameworks aimed at optimising decisions
made during this process. Section 1.2 provides an overview of the biopharmaceutical drug development process that includes probabilities of success, economics, and timelines for development. Section 1.3 reviews pertinent aspects of biopharmaceutical manufacture which includes capital cost requirements, cost of goods figures, and implications for drug pricing. Section 1.4 explores outsourced and partnered approaches as alternative options to complete in-house drug development and manufacture. Section 1.5 highlights computational frameworks and approaches available in literature that are relevant to this work. Finally, the aims and organisation of this thesis are discussed in section 1.6.

1.2. BIOPHARMACEUTICAL DRUG DEVELOPMENT

Biopharmaceuticals are typically protein-based therapeutics agents produced by modern biotechnological techniques (Walsh, 2005; Schellekens, 2002), and have been described as described as therapeutics of the twenty-first century Tsuji (2007). Examples include hormones (e.g. insulin for diabetes), growth factors (e.g. erythropoietin for anaemia), and monoclonal antibodies (e.g. infliximab for rheumatoid arthritis). These have been produced in either genetically engineered microorganisms (e.g. E.coli) or animal cell-lines (e.g. Chinese hamster ovary cells). This group of therapeutics represents a noteworthy progression in the development of treatments for conditions and diseases whose counteraction requires the molecular targeting specificity that protein based drugs have the capacity to provide. This thesis takes a particular perspective on addressing monoclonal antibody based therapeutics whose development, like all biopharmaceuticals, typically carries a higher risk of failure than success, requires significant expenditure in capital and operating costs, and is considerably lengthy. The development route for a biopharmaceutical is exhibited in Figure 1.1 and as it can be seen several stages of testing are required before a therapeutic can be approved for market.

Figure 1.1. The drug development process.
Target identification is the stage where at least one disease target is identified with the complementary screening of therapeutic compounds that can demonstrate some acceptable level of potential in counteracting disease. Recently, techniques to facilitate this process have included combinatorial chemistry which can generate whole classes of compounds quickly, genomics which allows a greater understanding of disease targets at cellular and molecular levels, and high throughput screening which uses robotics to screen and obtain data on the potency of compounds (Monane, 1998). Preclinical trials will test any promising therapeutics in animal models so as to study safety and toxicity. Patents are typically applied for upon successful completion of this period. Also, drug developers will apply to their respective regulatory body for legal permission to progress with testing on human subjects. In the USA this body is the Food and Drug Administration (FDA). In the UK the equivalent for this purpose is the European Medicines Agency (EMEA). Three stages of clinical testing ensue, with each stage requiring a larger number of human subjects than the previous. For phase I, II, and III clinical trials these numbers are typically and respectively 20-80, 100-500, and 1000-5000 (Pharmaceutical Research and Manufacturers of America, 2005). Phase I testing is aimed at determining safety and dosage information. Phase II clinical testing is used to demonstrate that the therapeutic carries a clinical efficacy that is statistically significant. Phase III trials are used to provide information for package labelling, such as side effects, and to make comparison to standard drugs used for treatment, if any exist. Upon successful completion of phase III clinical trials a drug developer will apply for a biologics license application (BLA), or the equivalent, for approval from the regulatory body to allow marketing of the therapeutic. Post approval, the drug developer is required to continue testing the marketed drug to study longer term effects on patients and to glean a detailed understanding of dosage requirements (Liang, 2002). This post approval testing is often referred to phase IV testing.

1.2.1. Probabilities of success

Biopharmaceutical therapeutics are typically subject to high failure rates when undergoing the development process. This is in part due to the highly structured nature of the approval procedure, and the stringent regulatory standards that exist for each stage of progression in the development process. Data for the transitional
probabilities of success of selected biopharmaceuticals for each stage in the development route from phase I clinical trials onwards is presented in Table 1.1. Given that success rates from phase I clinical trials to market are typically are in the range of 10%-35%, these probabilities typically result in a development route that is more likely to fail than succeed for any one drug. For certain products, particularly monoclonal antibodies, Table 1.1 shows that phase II clinical trials carry the smallest likelihood of success. Literature explaining why this feature in clinical trial success rates exists has not been found. A survey conducted by Pavlou and Reichert (2004) indicates that for some groups of recombinant protein drugs the likelihoods of successfully traversing the entire clinical trial process and gaining approval for marketing have increased, though they do not provide a specific reason for this. These groups are diabetes and endocrinology, and anti-infectives. It is also seen in Table 1.1 that aside from anti-neoplastic (anti-tumour antibiotics) therapeutics, the monoclonal antibody products generally carry a lower probability of success than other recombinant protein products. Of the monoclonal antibody products shown chimeric products exhibit a stronger likelihood of progressing through clinical trials than humanised products. Reichert et al. (2005) comment that their research suggests the possibility that more stringent selection procedures were applied with chimeric products, particularly with the selection of drugs entering phase III clinical trials where a transitional probability of 1 is observed. Overall, it should be underscored that the number of drugs explored in the studies mentioned are relatively small and thus the probabilities of success observed are likely to be subject to some variance as more biopharmaceuticals enter the clinical trial process. Such data is relevant from a portfolio management perspective because intuitively a major factor in the selection of drugs for a research and development portfolio is the probability of success expected from a particular drug candidate. Rates of success for individual development phases are also likely to influence decisions that require the allocation of significant capital. For example, a drug developer may consider whether it is more appropriate to outsource commercial manufacturing instead of building in-house manufacturing capacity for a particular drug candidate if it is likely to fail during phase III trials.
Table 1.1. Published clinical trial transition probabilities for biopharmaceutical therapeutics.

<table>
<thead>
<tr>
<th>Type</th>
<th>Application</th>
<th>n</th>
<th>Phase I to Phase II</th>
<th>Phase II to Phase III</th>
<th>Phase III to Market</th>
<th>Phase I to Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant protein*</td>
<td>D&amp;E</td>
<td>22</td>
<td>0.91</td>
<td>0.84</td>
<td>0.81</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>AIID</td>
<td>15</td>
<td>1.00</td>
<td>0.73</td>
<td>0.63</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>CV/haemostasis</td>
<td>35</td>
<td>0.88</td>
<td>0.75</td>
<td>0.39</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Anti-infectives</td>
<td>12</td>
<td>0.83</td>
<td>0.50</td>
<td>1.00</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Anti-neoplastic</td>
<td>9</td>
<td>1.00</td>
<td>0.13</td>
<td>1.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Chimeric MAb</td>
<td>All</td>
<td>39</td>
<td>0.69</td>
<td>0.36</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Oncological</td>
<td>21</td>
<td>0.60</td>
<td>0.75</td>
<td>1.00</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Immunological</td>
<td>9</td>
<td>0.67</td>
<td>0.36</td>
<td>1.00</td>
<td>0.24</td>
</tr>
<tr>
<td>Humanised MAb</td>
<td>All</td>
<td>102</td>
<td>0.80</td>
<td>0.51</td>
<td>0.69</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Oncological</td>
<td>46</td>
<td>0.76</td>
<td>0.60</td>
<td>0.80</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Immunological</td>
<td>34</td>
<td>0.88</td>
<td>0.43</td>
<td>0.80</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Source: a. Pavlou and Reichert (2004), and b. Reichert et al. (2005). Notation: n – number of drugs included in the survey, MAb – monoclonal antibody, D&E – diabetes and endocrine, AIID – arthritis, inflammation, and immune disorders, and CV – cardiovascular. Chimeric products are antibodies where variable regions are murine and constant regions are human. Humanised products are constructed with mouse antigen binding regions derived and the remainder derived from a human source.

On the subject of controlling probabilities of success during clinical trials, Dutton (2007) suggests that these can be maximised with intelligent design and preparation of clinical trials throughout. This includes commencing clinical trials only when enough information is known about the drug compound, working with regulatory agencies before clinical trials take place, and diligent use of preclinical predictive models. Whilst such information is important for managing drug development effectively it is difficult to include in portfolio management tool considerations as the impact on probabilities of success is difficult to quantify. In many cases the decision maker can only account for appropriate and evidenced probabilities of success even if it is postulated that a superior approach to typical industry practices exist. From an alternative perspective it is critically important that upon the selection of a portfolio management strategy, project execution fully reflects the assumptions made on performance factors such as the probabilities of clinical trial successes. This means that in some capacity such assumptions are used more as benchmarks or targets than as predictive values.
1.2.2. DEVELOPMENT ECONOMICS

Recent figures for the cost of development of a single drug have been indicated to be in excess of $800 million. DiMasi et al. (2003) have reported calculations that put this figure at $802MM. Adams and Brantner (2006) repeated this study and have reported their estimate to be $864MM. Boston Consulting Group (2001) reports their estimate to be $880MM. As an example of how these costs are allocated within the drug development process, sample cost data from DiMasi et al. (2003) are presented in Table 1.2. As expected, these costs increase considerably with each stage and the cost of phase III trials is nearly six times the cost of phase I trials. Interestingly, the data shows that the standard deviation for cost of development for each stage is comparable in magnitude to the mean and median values. With such variance, the meaning for drug developers taking a conservative perspective is that they can justifiably expect to spend double the mean and median estimates.

Table 1.2. Cost data for clinical trials.

<table>
<thead>
<tr>
<th>Clinical Testing Phase</th>
<th>Mean Cost</th>
<th>Median Cost</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>$15.2MM</td>
<td>$13.9MM</td>
<td>$12.8MM</td>
</tr>
<tr>
<td>Phase II</td>
<td>$23.5MM</td>
<td>$17.0MM</td>
<td>$22.1MM</td>
</tr>
<tr>
<td>Phase III</td>
<td>$86.3MM</td>
<td>$62.0MM</td>
<td>$60.6MM</td>
</tr>
</tbody>
</table>

Source: DiMasi et al. (2003)

It is clear that the costs in Table 1.2 do not sum to the cost figures stated in literature that are in excess of $800MM. This is because the above figures stated in literature also account for the cost of failed drug development projects. DiMasi et al. (2003) and Adams and Brantner (2006) use the same calculation method but different databases, which are the Tufts Center for the Study of Drug Development database and the Pharmaprojects database, respectively. Broadly, the average cost of progressing a drug through the approval stages is determined, which is then divided by the probability of obtaining marketing approval for a drug in phase I, and adjusted to reflect time value of money relative to when development commenced. The resulting figure is the total cost incurred in developing and approving one drug.
1.2.3. DEVELOPMENT TIMELINES

The average development time for a biopharmaceutical product has been reported to be as much as 12 years but is still consistently shorter than development times for their pharmaceutical counterparts (Ashton, 2001). Typical time to market figures for biopharmaceutical drugs are in the range of 6 to 10 years (Foo et al., 2001; Ashton, 2001; Werner, 2004; Reichert, 2003). Since 1970, the development time has exhibited a trend of increase over the long-term but more recently, since 2001, development times have become increasingly shorter (Reichert, 2003). Data published from the study conducted by Reichert (2003) is presented in Table 1.3 and show that development times for the therapeutics included in the study vary from 5.8 to 9.4 years. There appear to be no obvious trends that are specific to either recombinant proteins or monoclonal antibody therapeutics. For immunological applications monoclonal antibodies have shorter development times than recombinant therapeutics and the reverse is true for anti-neoplastic applications.

Table 1.3. Published development time data for biopharmaceuticals (years) developed during 1982-2001.

<table>
<thead>
<tr>
<th>Type</th>
<th>Application</th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
<th>Mean</th>
<th>Median</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant protein</td>
<td>Endocrine</td>
<td>10</td>
<td>4.3</td>
<td>4.7</td>
<td>1.4</td>
<td>1.1</td>
<td>5.8</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>Immunological</td>
<td>4</td>
<td>7.7</td>
<td>8.0</td>
<td>1.1</td>
<td>1.0</td>
<td>8.8</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Cardiovascular</td>
<td>4</td>
<td>4.6</td>
<td>5.0</td>
<td>1.8</td>
<td>1.4</td>
<td>6.4</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Anti-infective</td>
<td>3</td>
<td>8.2</td>
<td>7.1</td>
<td>1.2</td>
<td>1.1</td>
<td>9.4</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>Anti-neoplastic</td>
<td>4</td>
<td>3.9</td>
<td>4.3</td>
<td>2.3</td>
<td>2.3</td>
<td>6.2</td>
<td>5.9</td>
</tr>
<tr>
<td>Monoclonal antibody</td>
<td>Immunological</td>
<td>4</td>
<td>5.5</td>
<td>5.7</td>
<td>1.0</td>
<td>0.6</td>
<td>6.4</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>Anti-neoplastic</td>
<td>4</td>
<td>6.7</td>
<td>5.5</td>
<td>0.8</td>
<td>0.7</td>
<td>7.5</td>
<td>5.9</td>
</tr>
</tbody>
</table>

Source: Reichert (2003). Notation: n = number of drugs included in study.

1.3. BIOPHARMACEUTICAL MANUFACTURE

Biopharmaceutical manufacture is a complex and delicate process that is typified by high capital costs for the building of facilities and necessary equipment, and high associated operating costs. Biopharmaceutical products are also among the most technically complex to manufacture of any industry (Ransohoff, 2004), take longer to manufacture than chemical entities, and require production operations that are more
difficult to control (McGurk, 2004). In recent years, it has been reported that the biopharmaceutical industry has been faced with potential shortages in manufacturing capacity with decreasing revenue potential. This is also exacerbated by higher pressures from regulatory bodies (Pisano, 1997; Molowa, 2001). Although anticipated shortages may not have been realised (Ginsberg et al., 2002) the importance of this issue to the biopharmaceutical industry was clearly displayed in the general level of concern generated at the time. Consequently and critically, this demonstrates a link between a drug developer’s manufacturing operations and its long-term business strategy. Published studies have indicated this in a more generalised context, for example see Brown (1998) and Demeter (2003). Brown (1998) provides evidence that manufacturing strategy is a principal factor in the success of new product development by linking it to operational capabilities that can directly guide the direction and scope of a business. Demeter (2003) surveys over 700 companies and notes that return on sales figures are significantly higher for companies possessing a manufacturing strategy than for those who do not. For the biopharmaceutical industry in particular, aligning decisions with a business strategy forms a major factor in the decision of whether to make products in-house or to outsource their manufacture to a contractor (Ransohoff, 2004; Seymour and Galliher, 2002). Because of the significant costs and resources required, the manufacturing strategy undertaken is influenced by the approach to portfolio management and drug development issues. The size of the portfolio, success rates through clinical trials, doses and expected market demand for each drug dictate the potential production requirements. The budget, cost of drug development, and speed to market required constrain the amount of capital that can be allocated to manufacturing within a specific time-period and how long the company has to acquire manufacturing capacity. All of these factors increase the difficulty facing biopharmaceutical firms, especially smaller ones, in achieving success in the industry.

One driver of both capital and operating costs is that products such as monoclonal antibodies are usually required in high doses (Toi et al., 2004; Finesilver, 2003; Costello and Halverson, 2003) and emphasises the demand for the construction of facilities with high production capacities. This is further exacerbated by the increasing number of applications for biopharmaceuticals such as monoclonal antibodies (Lotgenberg, 2007, Reichert and Dewitz, 2006) which will require further
increases in capacity for the manufacture of complex biologics. Reichert and Dewitz (2006) have reported that between January 2000 and June 2005 more than 130 monoclonal antibodies entered clinical trials, and have noted the possibility that this number could reach 240 by 2010. Elsewhere, it has even been predicted that a shortage in manufacturing capacity would currently exist by as much as a quarter of demand (Molowa, 2001). More recently it has been noted that this expected shortfall did not materialise, rather it has actually been suggested that the industry currently holds an excess of manufacturing capacity (Thiel, 2004). Miller (2008) supports this view by reporting that in 2007 investments in biopharmaceutical manufacturing and laboratory facilities by the 15 largest biopharmaceutical corporations totalled $21.9B, their highest since 2004. A survey conducted by Langer (2007a) provides further evidence of excess manufacturing capacity by reporting figures for the utilisation of full manufacturing capacity in industry. For mammalian expression systems utilisation during 2003, 2005, and 2007 was 76%, 69%, and 64% respectively. For bacterial expression systems these figures were 71%, 61%, and 62% respectively. Both sets of figures show a general decrease in capacity utilisation since 2003. The main reasons provided given for this by Langer (2007a) include improving titres, whole process yields, relationships with CMOs, and managerial competency. From a portfolio management perspective such drivers imply that integrating manufacturing capacity is becoming more attractive due to increasing productivity, finding contract manufacturing capacity may be easier than in previous years, and perhaps the gap between planning and executing strategies is becoming narrower.

Manufacturing processes for biopharmaceutical products will usually consist of an animal cell culture (using CHO cells for example) or a microbial fermentation (using *E.coli* or *S.cerevisiae* for example) procedure that aims to first grow the host cell responsible for the production of the desired protein and then to express that protein in the desired quantity. This procedure is also known as upstream processing. Of the 31 therapeutic proteins approved between 2003 and 2006, 9 are produced in *E. coli* and 17 are produced by mammalian cell lines (Walsh, 2006). Mammalian cell culture is technically complex, slow and expensive, but these cell lines are used due to their capability to produce large therapeutic proteins that require post-translational modification such as glycosylation. With specific regard to antibodies, mammalian cell culture manufacturing routes show greater promise for the production of whole
fully-human monoclonal antibodies whilst bacterial cell culture appears to be most favourable for antibody fragments (Chadd and Chamow, 2001). As the product is initially crude it must be purified through a series of purification operations, collectively known as downstream processing. At each stage some of the crude product is lost, so increasing the efficiency of production is a challenge that biopharmaceutical manufacturers across the industry are interested in meeting even if only for the sake of increasing profitability. Additionally, all facilities must meet strict guidelines that are certified to align with current good manufacturing practice (cGMP) and it is critical to note that this strictness is further supported by the premise that the process by which biopharmaceuticals are manufactured defines their function and activity (Griffiths, 2004). Hence, biopharmaceutical manufacturers are pressured to design the commercial manufacturing process correctly in the first instance. Alongside significant set-up and running costs as well as strict regulatory guidelines, there is considerable uncertainty that impacts manufacturing process decisions. These include uncertainties in the necessity for glycosylation (Wemer, 1999), the clinical development timeline, the efficiency of manufacturing operations, and the annual output of the required product (Wemer, 2004).

Risk is an irremovable obstacle for the drug developer and it should be considered in the process of planning for drug development. Mohs (2008) recognises that there are three principal risks that can lead to failure in drug development: target risk, clinical development risk, and market risk. Target risk poses an obstacle for the drug developer in finding a disease target such that administration of the drug leads to ineffective treatment of the disease. This risk is exacerbated by the biological complexity of the disease. Clinical development risks include the technical probabilities of success in bringing a drug to market alongside uncertainties in development costs and the time taken to achieve market approval. Nicholson and Latham (1994) note that even a month’s delay in gaining marketing approval can lead to many millions of dollars in lost sales revenue. Market related risks include significant uncertainty in forecasted drug prices partly due to lengthy lead development times which can impact profitability considerably. Also, determining expected levels of demand can introduce difficulty in planning the capacity design specifications for manufacturing. Ultimately, target uncertainties influence the difficulty of finding effective drugs, developmental uncertainties can terminate
development of a drug before a return can ever be realised, and market related uncertainties influence whether a marketed drug ever meets the revenue expectations of the developer. Brastow and Rice (2003) further consider uncertainties in manufacturing capabilities, such as whole process yields, and finding high quality outsourcing capacity, should this be relevant to the drug developer. Nicholson and Latham (1994) also mention government intervention in drug pricing and healthcare reforms as risks that can significantly influence profitability. When managing the impact of uncertainty it is of interest to biopharmaceutical companies to delay fixed investment costs till as late as possible, minimise development and production costs, and preserve the ability to meet demand when eventually reaching the marketplace. Hence, a major and ubiquitous problem confronting biopharmaceutical drug developers is how and when to best make and implement critical business decisions so that important rewards such as profitability are optimised. Yet more expensive, risky and complex is the development of a portfolio of drug candidates, especially under the limitation of constrained resources. Within the framework of a company’s pipeline, valuing one drug at a time is not sufficient and drug developers must consider the entire portfolio under technological and market uncertainty and resource constraints (Rogers et al., 2002). Here the developer must also make decisions to best construct the portfolio such as to optimise the management of any resource and reward related trade-offs that each drug development project may introduce.

Other options for acquiring manufacturing capacity exist, such as outsourcing and partnering, and will be discussed later in this chapter. It is already observed that building full manufacturing capacity gives the highest level of managerial control and intellectual property acquisition, but also requires the largest outlay of capital and carries the highest penalty in the event of failure. Because of the significant financial and operational impacts drug developers are forced to consider manufacturing strategies years in advance of construction and must ponder whether they should even be manufacturing the product themselves. It has been reported that building a commercial scale biopharmaceutical manufacturing facility can take at least three years (Ginsberg et al., 2002, Werner, 2004; Thiel, 2004). In light of this and depending on the drug developer’s circumstance, alternative options can be particularly attractive.
1.3.1. CAPITAL COST REQUIREMENTS

Two broad classes of cost are of interest to drug developers in biopharmaceuticals, which are fixed capital investment and operating costs. The total fixed capital investment required includes the facility and equipment. The operating cost will typically include the cost of manufacturing materials, utilities, and wages. In this section, reported figures for these costs will be reviewed.

Pavlotsky (2004) conducted a study to approximate facilities cost, in which more than 100 projects involving construction of a pharmaceutical facility between 1993 and 2003 were investigated. Facilities in this study were considered in three sizes, 100,000 sq.ft., 50,000 sq.ft., and 25,000 sq.ft and the findings are summarised in Table 1.4. From these figures it is immediately apparent that from this study the cost of constructing biopharmaceutical manufacturing facilities in the USA is more expensive than doing the same outside of the US. No explanation for this is given by Pavlotsky (2004).

<table>
<thead>
<tr>
<th>Gross Size</th>
<th>Net Size</th>
<th>Cost Type</th>
<th>USA Facilities</th>
<th>Non-USA Facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,000 sq.ft.</td>
<td>70,000 sq.ft.</td>
<td>Total Cost</td>
<td>$89MM</td>
<td>$64MM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (gross)</td>
<td>$893</td>
<td>$644</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (net)</td>
<td>$1258</td>
<td>$920</td>
</tr>
<tr>
<td>50,000 sq.ft.</td>
<td>30,000 sq.ft.</td>
<td>Total Cost</td>
<td>$43MM</td>
<td>$31MM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (gross)</td>
<td>$862</td>
<td>$622</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (net)</td>
<td>$1437</td>
<td>$1036</td>
</tr>
<tr>
<td>25,000 sq.ft.</td>
<td>14,000 sq.ft.</td>
<td>Total Cost</td>
<td>$25MM</td>
<td>$18MM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (gross)</td>
<td>$984</td>
<td>$712</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$/sq.ft. (net)</td>
<td>$1757</td>
<td>$1271</td>
</tr>
</tbody>
</table>


Pavlotsky (2004) also looked to identify patterns in the cost of constructing biopharmaceutical manufacturing facilities (Table 1.5). It can be seen that the base building cost and the equipment cost together comprise in the region of 50% of the total facility cost. Comparing to Lang factors (Lang, 1948) which linearly correlates the capital cost of equipment to the total capital cost of the facility, the figure seen in Table 1.5 for process equipment cost appears somewhat excessive. This figure would
convert via its reciprocal to an equivalent Lang factor in the range of 3.3-3.7, whereas typical Lang factors for biopharmaceutical facilities have been considered to be from 4 to over 8 times the equipment cost (Novais et al., 2001). This suggests that process equipment costs should comprise 13-25% of the total facility cost. It should be taken into account that such numbers should be treated as general guidelines and that it is the specific design of the facility that determines the appropriate Lang factor.

**Table 1.5.** Breakdown of construction costs.

<table>
<thead>
<tr>
<th>Component</th>
<th>Cost (% of total facility cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base building cost</td>
<td>22-25%</td>
</tr>
<tr>
<td>Process equipment cost</td>
<td>27-30%</td>
</tr>
<tr>
<td>Engineering, management and administration</td>
<td>13-16%</td>
</tr>
<tr>
<td>HVAC systems</td>
<td>9-12%</td>
</tr>
<tr>
<td>Process piping and utilities</td>
<td>8-11%</td>
</tr>
</tbody>
</table>


Additionally, Pavlotsky (2004) investigated factors influencing the cost of the manufacturing facility (Table 1.6). The most influential factor amongst these was the competence in constructing the facility, which can influence the cost of construction by up to 15%. The terms 'luck of competence' or 'luck of ability to make decisions' were not defined further in the publication. In context, these terms are respectively taken here to mean the chance that the skills required to complete construction are indeed possessed, and that difficult problems were met with at least satisfactory solutions.
Table 1.6. Variables that influence the pharmaceutical cGMP facility construction cost (Pavlotsky, 2004).

<table>
<thead>
<tr>
<th>Component</th>
<th>Variation (±)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luck of competence and luck of ability to make decisions</td>
<td>±15%</td>
</tr>
<tr>
<td>Uniqueness of process equipment, single source suppliers or competitive bidding</td>
<td>±7%</td>
</tr>
<tr>
<td>Materials of construction</td>
<td>±5%</td>
</tr>
<tr>
<td>Utilities availability, requirements and constraints</td>
<td>±4%</td>
</tr>
<tr>
<td>Construction labour costs</td>
<td>±3%</td>
</tr>
<tr>
<td>Time-factor market constraints</td>
<td>±3%</td>
</tr>
<tr>
<td>Architectural concept, people movement and material flow organisation, open process area or multiple compartments organisation</td>
<td>±2.5%</td>
</tr>
<tr>
<td>HVAC requirements: air changes, unidirectional or turbulent airflow, and air pressure cascading, once-through air systems, active or passive differential pressure control</td>
<td>±2%</td>
</tr>
<tr>
<td>Materials supply just-in-time or in-place warehousing</td>
<td>±2%</td>
</tr>
<tr>
<td>Interpretation of validation or data acquisition requirements</td>
<td>±1%</td>
</tr>
<tr>
<td>Interpretation of life safety, fire safety, environmental safety, product safety and abatement requirements add or subtract 1%</td>
<td>±1%</td>
</tr>
</tbody>
</table>

Source: Adapted from Pavlotsky (2004)

It is clear that from Table 1.4 Pavlotsky (2004) correlates the floor size of the facility with the capital cost. An alternative approach of estimating cost might be based on the required output of therapeutic per year from the facility whilst accounting for additional technical manufacturing capabilities. Calculations from Werner (2004) with such an approach are displayed in Table 1.7. This clearly shows that significant cost savings in both capital expenditure and the cost of goods can result from increasing either the titre by which the therapeutic protein is expressed or the production yield. This highlights the importance of including such manufacturing characteristics into the economics calculations of biopharmaceutical manufacture. Werner (2004) mentions that cost savings can also result from shortening the fermentation process time. This is intuitive because for a constant demand this means that more batches would be produced per year but requiring lower process volumes and smaller equipment requirements.
Table 1.7. Facility cost calculations adapted from Werner (2004).

<table>
<thead>
<tr>
<th></th>
<th>0.1g/L</th>
<th>1g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fermentation titre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td>40%</td>
<td>70%</td>
</tr>
<tr>
<td>Capacity required</td>
<td>$62\times10^6$</td>
<td>$3\times10^6$</td>
</tr>
<tr>
<td>Number of bioreactors</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>Capital cost</td>
<td>$1,600$MM</td>
<td>$100$MM</td>
</tr>
<tr>
<td>Cost of goods per gram</td>
<td>$1500$</td>
<td>$260$</td>
</tr>
</tbody>
</table>

Assumptions: 10,000L fermenters, 250kg/year biopharmaceutical therapeutic required, and $50K cost per bioreactor.

Surprisingly, considering the significance of the topic to the industry there are few studies in peer reviewed literature available on the economics of biopharmaceutical manufacture. It has even been reported that capital requirements are underestimated (Kessell and Frank, 2007).

1.3.2. COST OF GOODS

Equally as important as the capital expense incurred in building a manufacturing facility is the cost associated with producing the product itself, otherwise known as the cost of goods. Studies on the cost of goods appear to be even lesser in number than capital estimates of biopharmaceutical manufacture despite it being one of the main drivers in the pricing of drugs (de Nohrona Pissarra, 2004). The principal cost driver in producing monoclonal antibodies is the bioreactor titre, even though chromatography may account for two-thirds of the downstream processing costs (Meyers, 2000). According to Meyers (2000), costs for monoclonal antibody production are approximately equally split between cell culture, purification, and support.

In a study published by Datar (1993) on tissue plasminogen activator (tPA) manufacturing, the operating costs and the distribution of the costs were analysed for mammalian and bacterial processes. An analysis of the annual production costs revealed that the annual direct manufacturing expense for the mammalian process was $70.9 million, approximately 60% of the total expense while that for bacterial $113
million, approximately 47%. A detailed breakdown constructed by the authors is shown in Table 1.8.

### Table 1.8. Annual operating expenses for the production of tPA.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Mammalian production</th>
<th>Bacterial production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>%</td>
</tr>
<tr>
<td>Fermentation</td>
<td>$23,405,910</td>
<td>33%</td>
</tr>
<tr>
<td>Recovery</td>
<td>$7,801,970</td>
<td>11%</td>
</tr>
<tr>
<td>Waste</td>
<td>$7,092,700</td>
<td>10%</td>
</tr>
<tr>
<td>Utilities</td>
<td>$12,057,590</td>
<td>17%</td>
</tr>
<tr>
<td>Labour and supervision</td>
<td>$6,383,430</td>
<td>9%</td>
</tr>
<tr>
<td>Pat./Royal.</td>
<td>$12,057,590</td>
<td>17%</td>
</tr>
<tr>
<td>Other</td>
<td>$2,127,810</td>
<td>3%</td>
</tr>
<tr>
<td>Total</td>
<td>$70,927,000</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: Datar (1993).

The analysis revealed that labour and supervision costs for the bacterial process were much higher than that for the mammalian process. This is most likely due to the complex recovery operations required following bacterial fermentations. The raw material costs were higher for mammalian processes, in part due to the expensive media required for the fermentation. One significant finding was the difference in the ratio between the cost of upstream processing and downstream processing for the two manufacturing routes. For the mammalian process this ratio was 3:1, while for the bacterial process it was 1:7, thereby highlighting the importance of the expression system in determining the distribution of costs within the cost of goods.

Similarly, Sommerfeld and Strube (2005) investigated the distribution of manufacturing expenses in the production of monoclonal antibodies (Table 1.9). In this study, a generic process flowsheet was established, based on existing monoclonal antibody production using mammalian cell culture methods in industry. To benchmark their findings they assume that the distribution of costs between the upstream and the downstream processes are in the range of 50-80% for downstream processes as suggested by Lowe (2001). Some important trends are demonstrated in Table 1.9 that are associated with an increase in fermentation titre. Within the distribution of the cost of goods attributable to downstream processing, equipment
becomes less of a driver, consumables become more important, and the total cost of goods attributable to downstream processing becomes more of a determinant for the total cost of goods.

**Table 1.9.** Influence of production rate on specific downstream costs.

<table>
<thead>
<tr>
<th>Downstream Cost Constituent</th>
<th>Product Titre Concentration (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1g/L</td>
</tr>
<tr>
<td>Equipment</td>
<td>65.7%</td>
</tr>
<tr>
<td>Consumables</td>
<td>16.8%</td>
</tr>
<tr>
<td>Waste</td>
<td>5.7%</td>
</tr>
<tr>
<td>Raw material</td>
<td>2.4%</td>
</tr>
<tr>
<td>Labour</td>
<td>8.1%</td>
</tr>
<tr>
<td>Other</td>
<td>1.3%</td>
</tr>
<tr>
<td>% of total cost of goods</td>
<td>45%</td>
</tr>
</tbody>
</table>

Source: Adapted from Sommerfield and Strube (2005).

### 1.3.3. Drug Pricing

Appropriate drug pricing is essential for profitability. It is difficult to find information on the pricing of biopharmaceutical products but an empirical investigation into the sales price of monoclonal antibody drugs yields an interesting finding (Figure 1.2). It is clear from Figure 1.2 that there is an inverse relationship between drug price and annual demand in that the higher the demand the cheaper the sales price per gram of therapeutic. This is intuitive as revenues of the drug must cover fixed capital costs of manufacturing facilities, the costs of development, and the costs incurred candidates that have failed during the clinical trial process. An extreme example is *Simulect*™ which at the time of investigation was produced at an approximate rate of 1 kg/year and commands a sales price of over $79,000 per gram.
1.4. OUTSOURCING AND PARTNERING

In-house development and production has been a traditional mode of operation for pharmaceutical and biopharmaceutical developers alike. In the earlier years of the biopharmaceutical industry patient and product demands were so high that production costs were of small concern, so the emphasis was simply on getting the therapeutic to the market to recoup the cost of development and generate profits (Kamarck, 2006). More recently, the need for alternative strategies has been emphasised in industry. One reason for this is that a biopharmaceutical manufacturing plant can cost $60MM to $600MM (Farid, 2006; Molowa, 2001; Werner, 2006). Kamarck (2006) comments that the typical capital cost of a biopharmaceutical facility is more than five times that of a chemical plant. Understandably it is difficult for start-up biopharmaceutical developers to finance the construction of such a facility. A second reason for this is that stringent regulatory guidelines for biologic products impose the consideration of building a facility years in advance of when it is needed. Importantly, building manufacturing facilities while still in the clinical trial process is unlikely to provide sufficient evidence to assure the developer that after their construction, the existence of some set of approved drugs will be awaiting manufacture. This is one of the reasons why building a biopharmaceutical manufacturing facility is perhaps regarded as the riskiest constituent of the drug development venture (DePalma, 2008). Additionally, overarching pressures such as long standing issues with research and
development productivity and investors demanding maintenance of profit growth (Belsey, 2007), stimulate the need to consider options outside of traditional drug development strategies.

Ultimately, existing alternatives to integrating all activities in-house can be appealing to drug developers as they allow for the risk and impact of failure to be mitigated as well as for expertise and productivity to be accessed expediently. Clinical development activities can be outsourced to a contract research organisation (CRO). Clinical and commercial manufacturing can be outsourced to a contract manufacturing organisation (CMO). Additionally, all of the capabilities offered by a CRO or CMO can potentially be offered by a single partner with the appropriate experience in the relevant technologies and the capacity to manage new projects. It is clear that the choice between any of these strategic routes is a key decision as biopharmaceutical companies must develop cost-effective development and manufacturing capabilities (Pavlou and Belsey, 2005) to assist in maintaining competitive advantages. Further, Langer (2007b) reports figures for organisations developing biopharmaceutical drugs fully in-house. For products manufactured via mammalian cell culture production in 2005 and 2006, 58% and 56% of organisations respectively conducted 100% of biopharmaceutical production in-house. For products manufactured using microbial expression systems these figures are 58% and 61% respectively. These figures indicate that biopharmaceutical developers have generally accepted alternative method to integrating manufacturing capacity in-house. The implications for portfolio management tools is that given this interest such tools should be capable of analysing what and when to outsource or out-license in order to meet manufacturing requirements.

1.4.1. CONTRACT RESEARCH ORGANISATIONS

CROs have the capacity to save development time and provide efficient access to expertise in the development of biopharmaceuticals. Additionally it is intuitive that a drug developer would be interested in limiting the requirements for significant upfront capital so that expenditures can be made and managed on a per-drug basis. Trends indicate that research and development costs incurred in bringing a drug to market have increased substantially with time. Undoubtedly this adds to the concern for
maintaining efficient cost and cashflow management. For example, historically in the US this cost has been $125MM in 1989, $231MM in 1993, and $450MM in 1997 (PhRMA, 1998). As mentioned, more recently this has increased to over $800MM. Outsourcing is not a new concept for the industry. In the 1970s, pharmaceutical industry outsourcing included clinical trials and more recently in the 1990s outsourcing came to include the early stages of target identification (Crossley, 2002). They are becoming increasingly important because of pressures on drug companies to develop new technologies in a changing regulatory environment and to move products to market sooner to gain competitive advantage in the marketplace (Maloff et al., 1997). In support of this view, DiMasi et al. (2003) report more recently that there is rapid growth in the outsourcing of research and development activities. Also, Liszewski (2007) reports that the CRO industry grew at an annual rate of 15% in 2007, where one significant reason cited for this growth is that outsourcing clinical trials is cheaper and probably more effective than managing them in-house. The growth in outsourcing is not just limited to activities that have had a firm establishment in the outsourcing industry. It is also observed that the scope on what can be outsourced is increasing significantly. Currently virtually all stages of drug development can be outsourced (Clark and Newton, 2004; Crossley, 2004). Thus biopharmaceutical developments are open to outsourcing as much or as little as they prefer. It has been reported in recent literature that outsourcing parts of the research and development effort has been a growing trend (Papadopolous, 2000), and recent figures indicate that of all research and development conducted in the pharmaceutical and biopharmaceutical industries 30-35% is channelled to CROs (Crossley, 2002; Clark and Newton, 2004). Additional advantages for use of a CRO include the freedom to concentrate on core competencies, greater flexibility, facilitation of rapid exploitation of technologies, outsourcing of weaker projects that yet have some reasonable likelihood of success, and access to a global body of expertise (Piachaud, 2002). An obvious concern about outsourcing is the risk of proprietary and commercially sensitive information being leaked to other companies. Such an event could substantially weaken a drug developer’s competitive advantage, especially if its pipeline is not technologically diverse. Understandably, it would be in the business interests of a CRO to guard against this because of disastrous repercussions. Piachaud (2002) finds that disadvantages of using a CRO include concerns about controlling the CRO, loss of critical internal skills, and difficulties in monitoring and evaluating the
performance of the CRO. Additionally, CROs receive payment through upfront and royalty charges on marketed products so whilst it is cost-effective in the short term to utilise a CRO, sacrificing a percentage of future revenues may prove to be more costly in the long term.

1.4.2. Contract manufacturing organisations

In recent years widespread concerns over the ability of industry capacity to meet the global demand for biopharmaceuticals has assisted in spurring acceleration in the building of manufacturing capacity by CMOs. The factors and concerns mentioned when outsourcing to a CRO are shared by those for a CMO. The rate of growth of the contract manufacturing industry has been estimated to be at 14% per annum (de Nohrona Pissarrah, 2004). Thiel (2004) reports that at the time of publication there were mixed views on the ability of the industry to keep pace with the production demand for biopharmaceuticals. The publication suggests that the general perception in the industry is that there is an excess in manufacturing capacity but some specialist consultants believe that this could be short-lived due to the expectation that increased market penetration of blockbuster biopharmaceuticals may reverse this situation by 2008. This excess is likely to be exacerbated by the finding that some biopharmaceutical drugs can be used in an expanded number of indications, or an increase in the rate of entry of drug candidates into clinical trials. Similar opinions held by biopharmaceutical developers may have spurred the use of CMOs at the time. Additionally, the impact of higher fermentation titres, especially those beyond 5g/L, for newer biopharmaceutical products entering development may improve the likelihood that industry-wide production capabilities are adequate for meeting demands.

As building a cGMP manufacturing capacity is capital intensive, outsourcing manufacturing responsibilities to a CMO reduces upfront fixed costs and converts them into flexible costs that are more manageable (Nähr and Nordström, 2005). Typically the cost per gram of manufacturing a biopharmaceutical with a CMO will be more expensive than with in-house production, and control over manufacturing activities will be compromised to some extent (Rajapakse et al., 2005). One principle motivation for using a CMO is that it is difficult to accurately forecast the demand
requirements for a particular therapeutic in the event that it reaches the marketplace. This carries the consequence that even if a drug successfully gains marketing approval the manufacturing facility can be significantly limited in its production capacity or be in excess of it. In such a circumstance the company is either limiting sales because of short manufacturing capacity or has spent too much on the facility and is incurring unnecessary operating and maintenance costs. It understandable that most small biopharmaceutical companies will choose to outsource manufacturing activities (Lias and Fogerty, 2002), however the option to utilise a CMO is not confined to smaller drug developers. Large biopharmaceutical companies are also using CMOs. For example, Genentech has outsourced some of its production requirements for Herceptin to Wyeth Biotech (Kamarck, 2006). Beyond concerns about manufacturing capacity lies the importance of access to skilled labour and the choice of expression systems, as the commercial scale production of biopharmaceuticals using mammalian cell culture is better understood than with using microbial systems (de Nohrona Pissarrah, 2004). One potential downside to using a CMO is that a contract must usually be secured much earlier than is needed (Nähr and Nordström, 2005). Like CROs, CMOs receive payment through upfront and royalty charges. Molowa (2001) has suggested that such royalty charges are in the region of 15% of sales revenue.

1.4.3. PARTNERSHIPS

Like the contracting industry for both CROs and CMOs, partnering is also becoming more popular amongst biopharmaceutical drug developers. In 2006 the value of alliances and mergers in the global biotechnology sector has been at its greatest since 2002, for which an increase in alliances with pharmaceutical companies was a principal factor (Lawrence, 2007). Partnering strategies came about in the industry as recombinant protein drugs were met with high demand and biopharmaceutical developers struggled to keep pace so partnering became a solution to reducing fixed capital costs by sharing manufacturing capacity (Kamarck, 2007). Recently, relatively few biopharmaceutical companies have even had the financial resources and in-house expertise in clinical development and marketing to bring a biopharmaceutical therapeutic to market without the use of a partner (Ashton, 2001). Partnering with a company capable of providing manufacturing capacity allows expenses, operational and financial risks, and knowledge to be shared (George et al.,
It is likely that the most important of these benefits is the prospect of sharing finances as this strategic option is essentially a source of finance. Also, it reduces the impact of failure and allows for wider product pipelines to be pursued which thereby assists in accelerating product development. Currently, pharmaceutical companies have strong interests in partnering with biopharmaceutical developers. This is because across the industry pharmaceutical companies have been widely reported to be lacking productivity in the development of their product pipelines (Drews, 2003; Scherer, 2007), and partnering with biotechnology companies is a solution that meets this need (Kessel and Frank, 2007). Pharmaceutical companies in this circumstance may also be under pressure to partner because at some point the patents on their products will expire, with severe associated losses in sales as generic versions of their products enter the market. Kessel and Frank (2007), also show that the aggregate annual value estimate of pharmaceutical licensing deals has accelerated dramatically since 2004 where it was under $30B, until 2006 where it was under $90B. However the publication also shows that the annual number of deals has only increased by a comparatively modest proportion, indicating involvement with much higher value projects, as is commonplace with biopharmaceutical research and development projects. Jones and Clifford (2005) confirm this in reporting that average values of a partnership deal has risen in recent years, being $110MM in 2003, $145MM in 2004, and $169MM in 2005. This trend is also supported by Jones (2007). The salient advantageous drivers that assist in forming partnerships appear to be shared by those for outsourcing. One important competency that large pharmaceutical companies possess which is not usually shared by contractors is in sales and marketing expertise alongside the access to established multi-national distribution networks. Downsides for the biopharmaceutical developer include that profits must be shared, and such alliances can be difficult to manage (Koza and Lewin, 2000). Importantly, the costs are very different especially when a marketed product results from the partnership. On the front end there are a variety of deals that the biopharmaceutical company can enter into that give biopharmaceutical developers a range of options for managing cashflow. However it is beyond the scope of this thesis to discuss these in detail. On the share of revenue, pharmaceutical companies appear to take a much larger proportion of revenue than contractors. Based on a series of interviews with 44 CEOs and business developers Moscho et al. (2000) indicates that large pharmaceutical partners will take in the region of 62% of revenues. 81% of this figure is rationalised
on the value of sales and marketing costs and expertise possessed by the partner. As these figures were ascertained through interviews with drug development chief executive officers it implies that industry sentiments agree that, in particular, it is difficult for smaller biopharmaceutical companies to replicate such competencies. Cunningham (2002) also discusses that smaller biopharmaceutical companies need access this type of expertise. 16% of the partner's revenue is justified through manufacturing costs and expertise. The remaining 3% is aimed at covering the value of previous clinical trial development. The paper also indicates that a considerable portion of this valuation is not derived from rigorously defined methodologies, and one-third of the interviewees admitted to this. Of this one-third, 21% add an arbitrary margin to their expected costs, and the remaining 12% derived the valuation of deal terms based on best guesses. These findings suggest that at the time of publication an industry wide valuation methodology was not in place which in itself may create the opportunity for the smaller biopharmaceutical companies to negotiate the deal terms within reason. In support, Fisher and Wang (2001) note that the most favourable offer that can be negotiated with a pharmaceutical partner is an equal split of the revenue.

Results from a study conducted by Danzon et al. (2005) which involved 900 pharmaceutical firms between the years 1988-2000 indicate another strong reason for partnering, which is that alliances have historically tended to increase clinical trial success probabilities particularly in phases II and III especially if the partner is large (Table 1.10). It should be noted that their data does not allow them to distinguish between the effects of strategic or opportunistic project selection and the genuine effect of alliances. They suggest that any negative effects of alliances in phase I may reflect the decision of developers who have recognised early that a drug candidate is already performing below expectations and have decided to out-license it to reduce the impact of failure, assuming that partners are undiscerning. Positive effects of alliances largely seen in later trials with larger partners may indicate that such partners are at least more competent in selecting good drug candidates to in-license or managing clinical trial processes at these stages. Alliances with small partners have a relatively small sample size and the negative effects seen in the dataset may be less robust. Danzon et al. (2005) also does not distinguish between effects concerning therapeutic protein and pharmaceutical candidates. Because of this it is difficult to use this data ubiquitously in modeling the impact of decisions made in the drug
Nevertheless this study indicates that it is possible for alliances to influence the probability of success for a drug development project.

Table 1.10. Effect of alliances on probability of advancing in a clinical trial.

<table>
<thead>
<tr>
<th>Size of original developer</th>
<th>Size of partner</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>None</td>
<td>4, 2</td>
<td>0, 4</td>
<td>2, 7</td>
</tr>
<tr>
<td>Small</td>
<td>-4, 4</td>
<td>7, 7</td>
<td>-7, 13</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>-9, 5</td>
<td>17, 4</td>
<td>9, 10</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-2, 4</td>
<td>13, 5</td>
<td>7, 9</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>None</td>
<td>4, 2</td>
<td>-3, 4</td>
<td>1, 7</td>
</tr>
<tr>
<td>Small</td>
<td>5, 4</td>
<td>-4, 10</td>
<td>-7, 14</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>2, 3</td>
<td>0, 7</td>
<td>15, 8</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-1, 3</td>
<td>12, 4</td>
<td>15, 6</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>None</td>
<td>-3, 1</td>
<td>-, -</td>
<td>-1, -</td>
</tr>
<tr>
<td>All sizes</td>
<td>3, 2</td>
<td>2, 4</td>
<td>11, 5</td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from Danzon et al. (2005). Note: In columns for phases I, II, and III the first number represents the marginal effect of the alliance and the second is the standard error, whilst ‘-’ means that the datum was not provided. The marginal effect is the additional impact on general probabilities of success, and is measured in percentage points. Small, medium, and large firms respectively considered to have 1 to 3, 4 to 24, and more than 25 drugs in development.

1.4.4. TIMING THIRD PARTY RELATIONSHIPS

After asking the question of when to include a third partner, the follow-on question entails which third party should be included. Literature that include some aspect of third party drug development tend to focus on the factors for doing so but do not usually discuss the stage at which third parties should be included. Arnold et al. (2002) looked at licensing deals in partnerships formed between large pharmaceutical companies and smaller biopharmaceutical companies. The study included a survey about the optimal time to commence a licensing deal. Their results showed that 60% of respondents believed phase II clinical trials to be the optimal time, 20% believed this to be phase I, and the remaining 20% believed this to be the phase III or later. They also analysed the value of licensing deals according to the stage of therapeutic development. Their results indicate that as drug candidates pass successive stages of clinical development there is a 22% increase in the value of the licensing deal. This preference for licensing during phase II is possibly due to most drugs failing at phase
II, so it is a means of gauging how successful the drug might be. Also, Arnold et al. (2002) comment that phase III trials are often too expensive for biopharmaceutical companies to conduct by themselves and in such cases the only choice is to license out the product. Intuitively, as drug candidates progress through clinical trials a reduced probability of failure is implied, so pharmaceutical companies are prepared to pay more money for further progression.

It should be considered that Arnold et al. (2002) researches licensing for a single drug and that possibly the decision on timing may change within a portfolio of partnerships, this is considered by Geyer et al. (1999). They stress that biopharmaceutical companies must strive for the correct portfolio of licensing deals as when time-wise cash flow characteristics are aggregated, distinctive cashflow profiles result for the company as a whole. The two studies mentioned in this section draw attention to the fact that timing the involvement of third parties is close to managing risks of product development and the expectation of cashflows. Clearly, different timing strategies have the capacity to produce different projected cashflows for the company. This is because different decisions made on the development of a drug will result in different payments being made at particular times. In the case of a licensing deal the biopharmaceutical company is open to different combinations of financing...
options that include signing the deal contract, royalty payments in the event that the therapeutic under contract upon market approval, milestone payments, and fees for services rendered by the third party. Also, in circumstances where it is solely the lack of capital that is restricting projected cashflows then any cash generated from previous licensing deals offers the option of covering the capital requirements for future projects. This would lead to a greater financial impact whether upon failure as well as upon successful approval of the therapeutic for marketing.

1.5. COMPUTATIONAL DECISION MAKING FRAMEWORKS

Computational approaches have been described in literature aimed at presenting a robust approach to optimising strategic decisions in pharmaceutical and biopharmaceutical development. Pertinent decisions that feature in such approaches include:

- The size of the drug development portfolio.
- The drug composition of the drug development portfolio.
- The schedule of associated developmental activities for selected drugs.
- In-house, outsourced, or partnered research and manufacturing.
- Facility and manufacturing process decisions.

Decisions made on the size and composition of the portfolio may also be influenced by imposing resource constraints. Scheduling of critical development activities subject to a prioritised order of development based on considerations such as the identification of the most promising projects and accounting for strategic windows of opportunity in targeted drug markets. The decision to keep development and manufacturing activities in-house may be subject to monetary constraints but the need to manage the impact of failure may be equally relevant. Ultimately, these decisions are explored by some algorithmic means to optimise at least one strategic criterion which is invariably a monetary goal such as net present value (NPV).

1.5.1. PORTFOLIO SELECTION AND CAPACITY PLANNING MODELS

Portfolio selection problems pertaining to drug development have been addressed in literature and cover a variety of methodological features, technical problems, and scenarios. Broadly, these approaches will either be based on the analysis of a limited
selection of strategic scenarios that are important to the drug developer, or based on
the specification of a decision making super-structure over which an optimal strategy
must be formulated by some strategic method of optimisation. Scenario analysis
based approaches can offer significant depth of analysis to highlighted scenarios but
optimisation-based approaches typically offer a greater capacity to realistically
represent the complexity of portfolio selection problems in drug development. Also,
optimisation-based approaches typically feature much larger decision spaces than
those presented in scenario analysis works and hence, can address solutions not
considered by such methods. Varma et al. (2008) present a method for optimising
expected net present value (ENPV) through strategic scheduling and resource
allocation for pharmaceutical research and development pipelines. The method
consists of a joint simulation and optimisation based framework that incorporates
mixed integer linear programming (MILP). Colvin and Maravelias (2008) describe a
multistage stochastic programming formulation for the strategic planning of clinical
trial scheduling in pharmaceutical research and development pipelines in order to
maximise ENPV. Subramanian et al. (2003) develop a simulation-based optimisation
framework capable of constructing research and development portfolios, ordering of
necessary project activities within the chosen portfolio, and dynamically reprioritising
activities based on real-time outcomes. They consider stochastic optimisation to
maximise mean net present value (NPV) and the probability of delivering a positive
NPV value, and their approach combines discrete-event simulation and MILP. Blau
et al. (2004) consider project selection and prioritising the order of development
based on maximising NPV and minimising the probability of achieving a negative
NPV using discrete-event simulation and a genetic algorithm based optimisation
procedure. They also consider the added complexity of a dependency structure
between different drug groups that pertain to capital cost, revenue, technical
probabilities of success, and experiential learning curves. Rogers et al. (2003) use
stochastic optimisation via MILP to determine the size and structure of
pharmaceutical drug development portfolios under budgetary constraints by using real
options valuation (ROV) to make continue and abandonment decisions. Other
relevant optimisation based approaches in this area include Choi et al. (2004),
Subramanian et al. (2001), Blau et al. (2000), and Jain and Grossmann (1999).
There are contributions that address the issue of capacity planning in the pharmaceutical industry using optimisation based approaches. For example, internal planning decisions of a company and the recognition of market opportunities have been considered by Gupta and Maranas (2004). Cheng et al. (2005) provide an approach to capacity planning and inventory control. A mathematical programming approach to capacity planning for the pharmaceutical industry was presented by Gatica et al. (2003). Levis and Papageorgiou (2004) provide a mathematical programming method for long-term, multi-site capacity planning under uncertainty for the pharmaceutical industry. Additionally, Maravelias and Grossman (2001) present a framework for the simultaneous optimisation of capacity planning and new product development. Other examples include Lakhdar et al. (2005) and Gatica et al. (2002). The majority of such approaches pertain to drug developers who have or plan to have integrated their development and manufacturing infrastructures in-house, however there are some who develop frameworks for wider settings. These can offer a more realistic setting as there are a significant number of biopharmaceutical companies that manufacture at least some of their drugs externally (Langer, 2004; Rogers, 2005). For example, Rogers and Maranas (2005) consider an approach for optimising the stage at which a drug developer should enter into a partnership licensing deal, the investment policy taken for each deal, and the portfolio structure of licensing deals to be held given a budgetary constraint using real option value (ROV) and mixed integer linear programming (MILP). Oh and Karimi (2004) develop a MILP technique for capacity expansion planning that considers decisions to expand existing manufacturing facilities, build new capacity, or outsource given regulatory factors. Rajapakse et al. (2005) present a framework based on scenario analysis and discrete-event simulation to evaluate scenarios for the developer to build manufacturing capacity early, build late, or outsource to a CMO. Further work in Rajapakse et al. (2006) provides some extension to this by using efficient frontier analysis to prioritise portfolio construction against resource constraints.

Approaches have been reported in the literature for the development of frameworks that incorporate both the problems of portfolio management and manufacturing capacity planning simultaneously. Intuitively, such approaches represent an important advancement while also presenting more challenging and realistic large-scale optimisation problems that offer vast extensions to the lines of inquiry for either
of the two problems in isolation. Levis and Papageorgiou (2004) develop a hierarchical approach based on MILP for maximising net present value (ENPV). The decisions to be optimised include product portfolio structure, the manufacturing network and investment strategy over multiple sites, design specifications for each manufacturing site, as well as sales and inventory plans. Maravelias and Grossman (2001) consider a MILP approach that the optimisation of NPV given decisions that include portfolio structure, sequencing and assigning resources to activities, outsourcing of activities, building new manufacturing capacity, and expanding existing manufacturing facilities. Their work features Lagrangean decomposition as a heuristic solution to the problem. Papageorgiou et al. (2001) present a MILP technique for optimising portfolio selection, proprietary manufacturing capacity and production plan design, schedule, and sales and inventory plans.

1.5.2. FACILITY AND MANUFACTURING PROCESS DECISIONS

Examples focused on facility and manufacturing process decisions in the biopharmaceutical industry have been featured in literature. Farid et al. (2001, 2005a) developed a tool to capture the technical, financial, and risk related aspects of competing biopharmaceutical manufacturing strategies. The tool combined process economics and logistics via discrete-event simulation, object-orientated programming, and uncertainties via Monte Carlo simulation. The work highlights a decision making framework based on multi-criteria decision making (MCDM) to reconcile pertinent financial attributes such as fixed capital investment, and non-financial attributes such as operational flexibility. Lim et al. (2005a, 2005b) further developed this approach to capture regulatory compliance activities as well as continuous production-related activities and applied it to evaluate the process economics of fed-batch or perfusion culture in mammalian cell culture processes. Mustafa et al. (2004) presents an approach for comparing the business impact of packed bed and expanded bed absorption chromatography operations.

1.5.3. STOCHASTIC OPTIMISATION APPROACHES

Stochastic models often offer a realistic representation of a given decision making environment. When developed for real situations such models result in large-scale
optimisation problems that can be difficult to solve. For problems featuring decision making under uncertainty there are two broad classes of solution scheme: multi-stage stochastic programming and stochastic optimal control (Cheng, 2005).

Stochastic programming usually involves two stages where the first stage accounts for strategic decisions made whilst the second stage accounts for the consequences of strategic decisions which manifest the performance of the strategy (Sahinidis, 2004). Optimisation-under-uncertainty problems are commonly formulated under such an approach where some form of linear programming such as MILP is also a popular solution scheme, as seen in section 1.5.1. Multi-stage stochastic programming extends this concept via the introduction of some set of procedures to enhance the efficiency of the optimisation approach.

Stochastic optimal control, or a Markov decision process, considers a model for sequential decision making that accounts for the outcomes of both current decisions and future decision making opportunities, (Cheng, 2003). Decision makers are intended to choose an action according to the state of the environment that they are currently in according to an optimal control policy. Dynamic programming techniques provide a means by which for this optimal control policy can be solved. For example, Cheng et al. (2003) utilises multiobjective stochastic dynamic programming as a solution scheme to optimal decision making for a chemical production plant.

Although both types of technique can be applied to the same problem but with different perspectives emphasised, Cheng et al. (2005) comments that stochastic programming is better suited for long-term strategic planning problems such as capacity planning. A well known barrier to either type of approach when using solution schemes that guarantee optimal solutions is the curse of dimensionality, which refers to the explosion in the number of possible combinations of decision variables as the number of decisions to be considered increases. Because of this a guaranteed optimal solution for some problems cannot be found as the problem has far outgrown current computational capabilities. In such cases heuristic approaches, which use some set of heuristics to reduce the decision space, are the next best option. Examples of heuristic approaches include simulated annealing (Kirkpatrick et al.
1983), TABU search (Glover and Laguna, 1997), genetic algorithms (Goldberg, 1989), and Lagrangean relaxation and decomposition (Fisher, 1981).

Strategic decision-making in the biopharmaceutical industry is considerably complex and represent large-scale optimisation problems, particularly when regarding strategies for research and development portfolio selection, scheduling, and the involvement of third parties. The types of problems that will be explored in this work will feature such decisions under a stochastic programming approach. Furthermore, features designed into the problem such as dependencies mean that problem is difficult to decompose because the performance of certain decisions is linked to other decisions. This is explained further in chapter 3. Due to the size and complexity of this problem heuristic techniques that aim to learn the structure of the problem are explored here. One reason for this is the anticipated computational efficiency offered by such approaches for such large-scale problems. A second reason is that such models have the capacity to give insight to the decision maker on the problem itself, which is rarely found in alternative methods.

1.5.4. HANDLING MULTIPLE CRITERIA AND OBJECTIVES

In problems involving capacity acquisition and expansion there may be many conflicting criteria to be considered for decision-making. The usefulness of extending the criteria considered was demonstrated by Farid et al. (2005b) who used multiple quantitative and qualitative criteria to assess the use of disposable components when building a biopharmaceutical manufacturing facility. This was supported by a framework that used multi-criteria decision making (MCDM). Additionally, Chhatre et al. (2007) presents a simulation-based approach to evaluate the impact of alternative process configurations under uncertainty using MCDM to amalgamate multiple criteria. Generally, the available types of methods for reconciling between multiple objectives include: transforming the multi-objective problem into a single objective problem, the lexicographic approach, and the Pareto approach. The field of MCDM offers numerous techniques that transform multi-objective problems into single-objective ones. This often involves assigning a weight to each objective so that relatively more important objectives contribute more to the output score of each alternative. In cases where a score is derived and in its simplest application, the most
preferred candidate solution will yield the optimal score. The lexicographic approach involves assignment of a preference order to each objective and then optimising each objective in order of preference until a non-dominated solution is found (for example see Sawik, 1997; and Volgenant, 2002). The Pareto approach (e.g. Cheng et al, 2003) compares each solution against each criterion and results in a set of Pareto optimal solutions that are said to be non-dominated. A solution is considered Pareto optimal if there is no other existing solution that can improve the value of one criterion without degradation in any other criteria. A key advantage of using lexicographic and Pareto approaches over transforming multiple objectives into a single objective is that they compare solutions according to each criterion. However, for numerous criteria it can be difficult for the decision maker to prioritise each individual objective, as in the lexicographic approach, without having a reasonable understanding of each one. Additionally, having a set of Pareto optimal solutions means that the decision maker still has the task of selecting a solution from this set. This can be especially difficult if there are more than two criteria as multiple trade-offs may exist between individual objectives. For these reasons the simpler technique of transforming multiple objectives into a single objective is focused on first in this work. Following this, exploration of a technique that includes Pareto optimal solutions will be explored.

The use of multiple criteria as a basis for analysing decisions can be supported effectively through MCDM which is one of the most well known branches of decision-making (Triantaphyllou, 2000). Its potential ability to support the selection of strategies in biopharmaceutical manufacturing lies in the flexibility it provides to reconcile both the financial and operational concerns of the decision maker. There are many methods available in MCDM which have been used to process financial and non-financial data. For example, Platts et al. (2002) used the weighted sum method (WSM) to analyse the decision to invest in internal manufacturing capabilities or to outsource these activities. Additionally, Steuer and Na (2003) have published a review of 265 publications that focus on utilising MCDM to aid decision-making in financial contexts.

1.6. AIMS AND ORGANISATION OF THESIS

An overview of the economics of biopharmaceutical development and manufacture has been presented alongside pertinent strategic decisions that must be made in this
process. In addition, published computational frameworks that have been developed for decision-support in this industry have been discussed. What is yet to be seen in literature or in the commercial software industry is the development a computational framework that unifies:

- Simulation of biopharmaceutical development and manufacture.
- Simulation-based estimation of the economics of biopharmaceutical development and manufacture.
- Optimal selection of therapeutics for a portfolio of a given size.
- Optimal order and year-by-year scheduling of drug development and manufacture such as to optimise the management of cashflows and inherent risks of failure.
- Optimal selection of CROs, CMOs, and partners if at all across the entire range of critical activities for the development and manufacture of therapeutics.
- Optimal year-by-year schedule for assignment of third party to critical development and manufacturing activities.
- Optimisation against multiple and possibly conflicting criteria.

Accordingly, the principal aim of this thesis is to develop a computational tool that is attentive to all of these features and incorporates them into a single framework. The ultimate aim of this work is the creation of a computational, intelligent, stochastic, combinatorial, and multi-objective decision-support framework that can optimise a strategic map for the year-by-year planning of critical development, manufacturing, and marketing activities for an entire portfolio of biopharmaceutical therapeutics, where key related findings in the development of this framework will also be distilled.

Chapter 2 develops the use of simulation and multi-criteria decision making (MCDM) as a means for gleaning and utilising diverse information about strategies for the manufacture for monoclonal antibodies that involve the use of third parties. The use of MCDM is inclusive of financial and operational criteria that are combined into a holistic framework that outputs a score. Use of the framework is illustrated in a scenario analysis example.

Chapter 3 develops a framework that unifies the concepts and strategic concerns described earlier in this section into a stochastic combinatorial multi-objective
optimisation framework. A super-structure for the formulation of a strategic development and manufacture plan is presented. The development and manufacturing environment is described. The optimisation algorithm is based on estimation of distribution algorithms (Mühlenbein and Paass, 1996) which are a recent development in the computer science discipline. An example case is formulated to demonstrate the use of the optimisation framework.

Chapter 4 extends the use of the framework presented in chapter 3 by exploring the impact of cashflow constraints and portfolio size on the algorithmic formulation of drug development and manufacturing strategies and their associated quality.

Chapter 5 explores the use of the manufacturing model for the development of straightforward equations for estimating key economics of biopharmaceutical manufacture within acceptable error margins. Equations are developed for both the manufacture of whole monoclonal antibodies in animal cell culture processes and antibody fragments in microbial processes. The equations are developed using multiple regression to learn geometric relationships between model outputs and inputs.

Chapter 6 explores the commercial considerations of this work and illustrates the economics of translating this work into a viable business venture. Chapter 7 discusses issues associated with validating this work. Chapter 8 provides the salient conclusions of this thesis and highlights avenues of investigation for extensions of this work. Finally, papers by the author published during the course of this work are attached in Appendix C.
CHAPTER 2

A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY

2.1. INTRODUCTION

This chapter investigates the design of a decision-support framework for the stochastic analysis of options for acquiring commercial-scale manufacturing capacity in a biopharmaceutical setting. It is recognised that acquiring biopharmaceutical manufacturing capacity is typically risky, expensive, lengthy, complex, and that relevant decisions are commonly based on a single metric, net present value (NPV). Hence, the overarching motivation behind this work is to explore whether the extension of decision-making criteria captured within a holistic framework can aid the decision maker towards making more informed choices regarding strategies for acquiring biopharmaceutical manufacturing capacity. The framework considers decisions faced by the drug developer and multiple criteria pertinent to decision-making in this setting. Decisions are mapped in a model that simulates their impact in the environment for capacity acquisition alongside salient uncertainties. One clear challenge is how to structure such a framework as different types of criteria may be relevant to the decision maker. Such criteria might be broadly categorised as financial, referring to the stability of the financial balance sheet and how effectively finances are used, or operational, broadly pertaining to implications of a more managerial nature. Within these two broad categories it is demonstrated that further categorisation exists. Because of these issues a multi-criteria decision-making (MCDM) technique is utilised here as a means for organising and reconciling multiple criteria. Hence, the overall decision-support framework combines a stochastic simulation model with a MCDM technique.
This chapter is organised as follows: section 2.2 details the architecture of the decision-support framework, section 2.3 describes the case study used to illustrate use of the framework, section 2.4 discusses the results, section 2.5 details the conclusions regarding the obtained results, and section 2.6 lists the nomenclature used in this work.

2.2. DECISION-SUPPORT FRAMEWORK

The decision-support framework was designed to model the financial and operational perspectives of strategies for acquiring commercial manufacturing capacity. Microsoft Excel was used for its implementation and an overview is presented in Figure 2.1. A hierarchical approach was applied to facilitate a clear and detailed representation of the business and manufacturing processes involved. This approach also offers the ability to rapidly change assumptions so that a wider variety of scenarios can be captured and analysed. The framework has four elements: a biomanufacturing process model, a profit and loss model, an MCDM technique, and a set of criteria used to distinguish between the strategic options.

---

**Figure 2.1.** An overview of the MCDM framework.
A decision-making scenario is comprised of a set of alternative options with each one representing a capacity acquisition strategy. Each option is defined as a collection of probability distributions to be inputted to the model. These inputs can be split into technical, commercial, business and qualitative categories. Technical inputs describing the manufacturing capabilities, such as the fermentation titre and the overall process yield, are directed to a biomanufacturing cost model. These technical inputs are used to generate data such as the cost of goods sold per gram (COGS) and the capital investment required. Commercial inputs include the market capture and the market lifetime of each drug. These inputs are used to define the production capacity required and expected sales revenue. Inputs from the business category define the structure of any contractual agreements such as royalty payments and the terms of any partnerships formed. Qualitative aspects of each option include the suitability of the location, control and flexibility over the projects considered, and the expected acquisition of manufacturing knowledge. Qualitative scoring is used to describe most of the operational aspects. The outputs of the biomanufacturing cost model are directed to the profit and loss model. The outputs of the profit and loss model determine the values of the financial criteria used to discriminate between the alternative options. Subjective operational inputs are fed directly to their corresponding criteria. Details of the biomanufacturing model and the profit and loss model will be discussed in further detail in sections 2.2.1 and 2.2.2 respectively.

2.2.1. BIOMANUFACTURING PROCESS MODEL

The framework used to model the manufacture of biopharmaceutical products (Figure 2.2) is based on work reported by Farid et al (2000, 2005a) and Lim et al (2005a and b). Figure 2.2 represents a typical manufacturing route for monoclonal antibodies. Included in the model are the main process and ancillary tasks involved in the manufacturing process as well as equations for calculating the utilisation of equipment, materials, utilities, and labour. The cost calculations are supported by a database of unit costs for equipment and materials used in the manufacturing process.
Figure 2.2. The biomanufacturing process model. Abbreviations: DSP - downstream processing, QC/QA - quality control / quality assurance, FCI - fixed capital investment, COGS - cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. The DSP yield is the overall yield after all downstream processing steps have been completed. The batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.
A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY

The overall input parameters to the model are the annual demand, the expected fermentation titre, the overall product yield and the anticipated success rate of each batch. Each unit operation has a process model comprising of design equations and mass balances. These are used to size equipment, determine the composition of the output streams and the amount of materials required (e.g. chromatography buffers). More details on these models can be found in Farid et al. (2007). Equipment sizes are determined by matching processing requirements such as volume to a database of equipment dimensions available at the time of writing. If the processing requirements require an equipment size that is in excess of available equipment dimensions then the sizes of equipment allocated will be such that all units will be of equal size with number clearly being sufficient to handle process requirements. The outputs from this model are the fixed capital investment for the manufacturing plant as well as the COGS. The fixed capital investment was estimated by the Lang Factor method (Lang, 1948) which correlates this figure to the equipment cost. The COGS model employed is shown in Table 2.1.

Table 2.1. COGS model breakdown.

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>Description</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Cost of Goods</td>
<td>Direct raw materials f(utilisation)</td>
<td>(2.1)</td>
</tr>
<tr>
<td></td>
<td>Miscellaneous materials 0.5 × direct raw materials</td>
<td>(2.2)</td>
</tr>
<tr>
<td></td>
<td>Direct utilities f(utilisation)</td>
<td>(2.3)</td>
</tr>
<tr>
<td></td>
<td>Operating labour f(utilisation)</td>
<td>(2.4)</td>
</tr>
<tr>
<td></td>
<td>Supervisors 0.2 × operating labour</td>
<td>(2.5)</td>
</tr>
<tr>
<td></td>
<td>Quality control and quality assurance 1 × operating labour</td>
<td>(2.6)</td>
</tr>
<tr>
<td></td>
<td>General management 1 × operating labour</td>
<td>(2.7)</td>
</tr>
<tr>
<td>Indirect Cost of Goods</td>
<td>Maintenance 0.1FY</td>
<td>(2.8)</td>
</tr>
<tr>
<td></td>
<td>Local taxes 0.02FY</td>
<td>(2.9)</td>
</tr>
<tr>
<td></td>
<td>Insurance 0.01FY</td>
<td>(2.10)</td>
</tr>
<tr>
<td></td>
<td>Capital charge F (1 + rate of interest)^C⁻¹</td>
<td>(2.11)</td>
</tr>
<tr>
<td></td>
<td>General utilities Cost per unit area per year × Facility size × Y</td>
<td>(2.12)</td>
</tr>
<tr>
<td>Total Cost of Goods</td>
<td>Direct Cost of Goods + Indirect Cost of Goods</td>
<td>(2.13)</td>
</tr>
<tr>
<td>Total Cost of Goods per Gram (COGS)</td>
<td>Total Cost of Goods / Annual Production Output</td>
<td>(2.14)</td>
</tr>
</tbody>
</table>

Note: Where F is the fixed capital investment, Y is the project duration in years, and C is the capital charge period. Source: Mustafa et al. (2004).
2.2.2. Profit and Loss Model

Table 2.2. Profit and loss model.

<table>
<thead>
<tr>
<th>Year</th>
<th>Year t</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income</strong></td>
<td></td>
</tr>
<tr>
<td>Dose per Patient per Year</td>
<td>$I_{1,t}$</td>
</tr>
<tr>
<td>Target Population Size</td>
<td>$I_{2,t}$</td>
</tr>
<tr>
<td>Market Capture, $I_{3,t}$</td>
<td>$f(t_0, M, t_0)$ (2.15)</td>
</tr>
<tr>
<td>Annual Product Demand, $I_{4,t}$</td>
<td>$I_{1,t}I_{2,t}I_{3,t}$ (2.16)</td>
</tr>
<tr>
<td>Estimated Sales Price*, $I_{5,t}$</td>
<td>59.832$I_{3,t}^{0.4672}$ (2.17)</td>
</tr>
<tr>
<td>Total Sales, $b_t$</td>
<td>$I_{6,t}$ (2.18)</td>
</tr>
<tr>
<td>Other Income</td>
<td>$I_{6,t}$ (2.19)</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td></td>
</tr>
<tr>
<td>Capital Investment</td>
<td>$E_{1,t}$</td>
</tr>
<tr>
<td>COGS</td>
<td>$v_t$ (2.20)</td>
</tr>
<tr>
<td>General and Administrative, $E_{3,t}$</td>
<td>$gI_{6,t}$ (2.20)</td>
</tr>
<tr>
<td>Royalties, $E_{4,t}$</td>
<td>$rI_{6,t}$ (2.21)</td>
</tr>
<tr>
<td>Total Plant Expenses to be Paid, $E_{4,t}$</td>
<td>$p_t(E_{1,t} + E_{2,t} + E_{3,t} + v_t)$ (2.22)</td>
</tr>
<tr>
<td>Profits owed to partner, $E_{5,t}$</td>
<td>$p_t(I_t - E_{4,t})$ (2.23)</td>
</tr>
<tr>
<td>Convertible Debt Repayments</td>
<td>$cL$ (2.24)</td>
</tr>
<tr>
<td><strong>Total Annual Expense, $E_t$</strong></td>
<td>$E_{3,t} + E_{6,t}$ (2.25)</td>
</tr>
<tr>
<td><strong>Profit</strong></td>
<td></td>
</tr>
<tr>
<td>Profit Before Interest and Tax, $P_{1,t}$</td>
<td>$I_t - E_t$ (2.26)</td>
</tr>
<tr>
<td>Depreciation</td>
<td>$P_{2,t}$</td>
</tr>
<tr>
<td>Taxable profit, $P_{3,t}$</td>
<td>$P_{1,t} - P_{2,t}$ (2.27)</td>
</tr>
<tr>
<td>Tax, $P_{4,t}$</td>
<td>$mP_{3,t}$ (2.28)</td>
</tr>
<tr>
<td><strong>Profit after Interest and Tax, $P_t$</strong></td>
<td>$E_t - (P_{2,t} + P_{4,t})$ (2.29)</td>
</tr>
<tr>
<td><strong>Net Present Value</strong></td>
<td></td>
</tr>
<tr>
<td>Annual Present Value, $P_{PV,t}$</td>
<td>$P_{PV,t}$ (2.30)</td>
</tr>
<tr>
<td>Net Present Value, $P_{NPV,t}$</td>
<td>$\sum_{t=1}^{T} P_{PV,t}$ (2.31)</td>
</tr>
<tr>
<td><strong>Expected Net Present Value</strong></td>
<td></td>
</tr>
<tr>
<td>Transition Probability</td>
<td>$\pi_t$</td>
</tr>
<tr>
<td>Cumulative Transition Probability, $\pi_c$</td>
<td>$\prod_{t=1}^{T} \pi_t$ (2.32)</td>
</tr>
<tr>
<td>Expected Net Present Value, $P_{ENPV}$</td>
<td>$P_{NPV,t} \pi_c$ (2.33)</td>
</tr>
</tbody>
</table>

Note: $t_0$ is the ramp period from commercialisation to peak market capture, $M$ is the peak market capture, and $t_f$ is decay period after peak market capture, $g$ is the proportion of sales that are expected to be spent on general and administrative costs, $r$ is the proportion of sales that will be used to pay royalty fees, $p_t$ is the proportion of expenses that the company is obligated to pay, $p$ is the proportion of profits it owes to any partners, $c$ is the coupon rate payable on the par value of the issued convertible debt $L$, and $m$ is the tax payable as a percentage of profit, and $a$ is the discount factor.

*Derived by correlating sales price versus demand for therapeutic monoclonal antibodies.
A profit and loss model was built to process income and expense data (Table 2.2) and to generate values for the decision-making criteria. It was found that the sales price data for monoclonal antibody therapeutics could be explained by a geometric regression model:

\[ P = 59.382D^{-0.4672} \] (2.34)

where \( P \) – sales price, \( D \) – annual demand of the therapeutic. The coefficient of determination \( R^2 = 0.8055 \) for this relationship meaning that over 80% of the data seen can be explained by the geometric model. In addition to the business revenues achieved from the sale of therapeutics other sources of finance are also modelled. These additional sources of finance include an initial public offering (IPO) and the issuance of convertible debt. It is assumed that this will be in the form of convertible bonds which are issued in the same year as the IPO. The holder of any convertible bonds can convert them to common stock of the issuing firm according to conditions specified by the firm prior to issuing the bonds (Kimura and Shinohara, 2006). The static conditions are the term of the bond, its coupon, the probability of conversion and the conversion premium which is the amount the bond holder must pay the firm in order to convert. The only variable condition is the year of conversion, providing this happens, which is modelled as a triangular distribution. Additionally, the year of conversion is expected to coincide with the successful commercialisation of the first drug which is reflected in the probability distribution. The amount of finance required to be raised through the issue of this instrument is calculated. The model assumes that there are no possibilities of negative cashflows. If the cashflow in any year for any scenario falls below $5 million then enough is raised to correct this with an upper limit of $10 million for the same year. It is assumed that there are no restrictions to issuing convertible debt at the time of need.

2.2.3. CRITERIA

The criteria used in the framework (Table 2.3) were chosen to represent a broad spectrum of values that may be important when considering detailed financial and operational perspectives. Also, each criterion is intended to capture a certain aspect of each option over its lifetime from an overall perspective. In the financial category
five categories of criteria are considered: profitability, asset utilisation, liquidity, long-
term solvency and capital structure. In the operational category four categories were
considered: productivity, the suitability of the location, control and flexibility over the
projects considered, and the expected acquisition of manufacturing knowledge. The
qualitative component of the productivity category includes the ease of training
manufacturing personnel. The location category includes the ability of the local
infrastructure to support operations and logistics, and the access to qualified
manufacturing personnel. The flexibility category includes the ability to control its
drug manufacture projects, the ease of expanding operations, and the ease of
consolidating manufacturing operations. The manufacturing knowledge category
includes the readily available manufacturing expertise, the number of company
personnel assigned to manufacturing, the potential to acquire manufacturing
knowledge, and the control over any manufacturing knowledge acquired.

When applying these criteria to a decision-making scenario a higher criterion value
represents a stronger position. Some of these criteria, such as those based on costs
had to be inverted to reflect stronger positions as higher values rather than as lower
values.
Table 2.3. Criteria used in the MCDM framework.

<table>
<thead>
<tr>
<th>Group</th>
<th>Category</th>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profitability</td>
<td></td>
<td>ENPV</td>
<td>$C_1 = \sum_{n=1}^{N} \left( \frac{\sum_{t=1}^{T} a_{t,n} P_{t,n}}{\sum_{n=1}^{N} a_{t,n}} \right)$ (2.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Profits to Total Sales</td>
<td>$C_2 = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Profits to Total Assets</td>
<td>$C_3 = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Profits to Total Equity</td>
<td>$C_4 = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.38)</td>
</tr>
<tr>
<td>Financial</td>
<td>Asset Utilisation</td>
<td>Total Sales to Total Fixed Assets</td>
<td>$C_5 = \sum_{n=1}^{N} \sum_{t=1}^{T} b_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Profit to Total Current Assets</td>
<td>$C_6 = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.40)</td>
</tr>
<tr>
<td></td>
<td>Liquidity</td>
<td>Total Current Assets to Total Current Liabilities</td>
<td>$C_7 = \sum_{n=1}^{N} \sum_{t=1}^{T} a_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.41)</td>
</tr>
<tr>
<td></td>
<td>Long-Term Solvency</td>
<td>Total Equity to Total Liabilities</td>
<td>$C_8 = \sum_{n=1}^{N} \sum_{t=1}^{T} a_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.42)</td>
</tr>
<tr>
<td></td>
<td>Capital Structure</td>
<td>Total Assets to Total Equity</td>
<td>$C_9 = \sum_{n=1}^{N} \sum_{t=1}^{T} a_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Sales to Total Equity</td>
<td>$C_{10} = \sum_{n=1}^{N} \sum_{t=1}^{T} b_{t,n} \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.44)</td>
</tr>
<tr>
<td>Operational</td>
<td>Productivity</td>
<td>Average Inverted COGS</td>
<td>$C_{11} = N \left[ \sum_{n=1}^{N} u_{n} \right]^{-1}$ (2.45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Profit to Total Manufacturing Personnel</td>
<td>$C_{12} = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \left[ \sum_{n=1}^{N} q_{n} \right]$ (2.46)</td>
</tr>
<tr>
<td></td>
<td>Qualitative Productivity Score</td>
<td>$C_{13} = \sum_{n=1}^{N} I \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Qualitative Location Score</td>
<td>$C_{14} = \sum_{n=1}^{N} I \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.48)</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td>Qualitative Flexible Score</td>
<td>$C_{15} = \sum_{n=1}^{N} I \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.49)</td>
</tr>
<tr>
<td></td>
<td>Manufacturing Knowledge</td>
<td>Qualitative Manufacturing Knowledge Score</td>
<td>$C_{16} = \sum_{n=1}^{N} I \left[ \sum_{n=1}^{N} a_{t,n} \right]^{-1}$ (2.50)</td>
</tr>
</tbody>
</table>

*Note:* For symbol definitions refer to the *Nomenclature* section.
2.2.4. THE WEIGHTED SUM METHOD

The weighted sum method (WSM) was chosen to evaluate the criteria shown in Table 2.3. These criteria are not all measured in the same dimensions and thus result in a multi-dimensional decision-making scenario. For comparisons between options to be drawn meaningfully and usefully the criteria values must be converted into an equivalent set of dimensionless numbers (e.g. Triantaphyllou, 2000). Normalisation can convert the criteria values as described and a normalised criterion value can be represented as:

\[ x_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}} \]  

where \( e_{ij} \) is the value of \( i^{th} \) criterion of the \( j^{th} \) alternative option and \( x_{ij} \) is the normalised rating of attribute \( i \) for the alternative option \( j \).

Each criterion is given a weighting that can be configured by the decision maker to reflect how important each criterion is in the decision-making process. These weight values must also be converted to an equivalent set of dimensionless numbers. The normalised weight value is:

\[ w_i = \frac{z_i}{\sum_{i=1}^{n} z_i} \]  

where \( z_i \) is the weight value of the \( i^{th} \) criterion.

The score generated by the WSM method can be represented as:

\[ S_j = \sum_{i=1}^{n} w_i x_{ij} \]
where $w_i$ is the normalised weight assigned to attribute $i$, and $x_j$ is the normalised rating of attribute $i$ for the alternative option $j$. The preferred alternative option has the highest score.

2.2.5. Uncertainty

To determine the stochastic ranking configuration, input values were specified as triangular distributions characterized by minimum, maximum, and most likely values. These inputs were subjected to a Monte Carlo simulation and the options were ranked according to their mean WSM score. Subsequent analyses were conducted to identify any tradeoffs linked to these mean scores.

2.3. Case Study Description

A hypothetical case study was formulated to illustrate and examine the use of the framework in capturing the financial and operational perspectives of decision-making scenarios in the acquisition of commercial scale biomanufacturing capacity. The example is based on a biopharmaceutical company that needs to acquire commercial manufacturing capacity and has three monoclonal antibody drug candidates in its product pipeline. One has just recently entered late stage clinical trials whereas the others are still in the early stages of clinical testing. The first, second and third drugs are required in medium, high, and low demands respectively. Given the progress of the drug portfolio toward commercialisation the necessity to acquire commercial manufacturing space is urgent. The company’s Initial Public Offering (IPO) has raised $100 million. The following possible options have been identified:

- **Partner’ option**: Partner with a large pharmaceutical company and split the expenses and profits.
- **‘CMO’ option**: Outsource all manufacturing requirements to a contract manufacturing organisation (CMO).
- **‘Build’ option**: Build a new facility and undertake all of the manufacturing by themselves.
- **‘Partner/Build’ option**: A hybrid option involving following through with the partnership but only for the first drug. The remaining drugs will be manufactured by the construction of a new facility.
A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY

- 'CMO/Build' option: A hybrid option involving a contract manufacturer but only for the first drug. The remaining drugs will be manufactured by the construction of a new facility.

Table 2.4. Market and development characteristics of drugs in the case study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose (grams per patient per year)</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Annual Market Size (patients)</td>
<td>100,000</td>
<td>200,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Clinical Trial Stage</td>
<td>Phase III</td>
<td>Phase I</td>
<td>Phase I</td>
</tr>
<tr>
<td>Expected Time to Market Entry (years)</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Expected Market Lifetime (years)</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2.5. Input values used for the case study.

<table>
<thead>
<tr>
<th>Input</th>
<th>Partner</th>
<th>CMO</th>
<th>Build</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected market capture – Drug 1 (%)</td>
<td>Tr(5.5, 11, 16.5)</td>
<td>Tr(5, 10, 15)</td>
<td>Tr(5, 10, 15)</td>
</tr>
<tr>
<td>Expected market capture – Drug 2 (%)</td>
<td>Tr(16.5, 33, 49.5)</td>
<td>Tr(15, 30, 45)</td>
<td>Tr(15, 30, 45)</td>
</tr>
<tr>
<td>Expected market capture – Drug 3 (%)</td>
<td>Tr(11, 22, 33)</td>
<td>Tr(10, 20, 30)</td>
<td>Tr(10, 20, 30)</td>
</tr>
<tr>
<td>Ramp time to peak market capture (years)</td>
<td>Tr(3, 2, 1)</td>
<td>Tr(4,3,2)</td>
<td>Tr(4,3,2)</td>
</tr>
<tr>
<td>Decay time from peak market capture (years)</td>
<td>Tr(3, 2, 1)</td>
<td>Tr(4,3,2)</td>
<td>Tr(4,3,2)</td>
</tr>
<tr>
<td>Transition probability – Phase 1 to 2</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Transition probability – Phase 2 to 3</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Transition probability – Phase 3 to FDA</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Review</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Discount Rate</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Fermentation titre (g/L)</td>
<td>Tr(0.6, 0.8, 1.2)</td>
<td>Tr(0.8, 1.0, 2.0)</td>
<td>Tr(0.4,0.6,1.0)</td>
</tr>
<tr>
<td>DSP yield (%)</td>
<td>Tr(55, 60, 65)</td>
<td>65</td>
<td>Tr(50, 55, 60)</td>
</tr>
<tr>
<td>CMO’s mark-up on COGS (%)</td>
<td>-</td>
<td>80</td>
<td>-</td>
</tr>
<tr>
<td>CMO’s royalty charge on sales (%)</td>
<td>-</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Profit paid to partner (%)</td>
<td>60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Expenses paid by partner (%)</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ease of training staff</td>
<td>Tr(5, 7, 9)</td>
<td>Tr(1, 1, 3)</td>
<td>Tr(3, 5, 7)</td>
</tr>
<tr>
<td>Ability of the local infrastructure to support operations and logistics.</td>
<td>Tr(6, 8, 10)</td>
<td>Tr(6, 8, 10)</td>
<td>Tr(8, 10, 10)</td>
</tr>
<tr>
<td>Ability to acquire manufacturing personnel</td>
<td>Tr(6, 7, 9)</td>
<td>Tr(8, 10, 10)</td>
<td>Tr(3, 5, 7)</td>
</tr>
<tr>
<td>Ability to control and manage drug manufacture</td>
<td>Tr(3, 5, 7)</td>
<td>Tr(8, 10, 10)</td>
<td>Tr(8, 10, 10)</td>
</tr>
<tr>
<td>Ease of manufacturing expansion</td>
<td>Tr(4, 5, 7)</td>
<td>Tr(5, 7, 9)</td>
<td>Tr(3, 5, 7)</td>
</tr>
<tr>
<td>Ease of manufacturing consolidation</td>
<td>Tr(1, 3, 5)</td>
<td>Tr(1, 1, 3)</td>
<td>Tr(8, 10, 10)</td>
</tr>
<tr>
<td>Readily available manufacturing expertise</td>
<td>Tr(5, 7, 9)</td>
<td>Tr(8, 10, 10)</td>
<td>Tr(1, 1, 3)</td>
</tr>
<tr>
<td>Potential to acquire manufacturing knowledge</td>
<td>Tr(5, 7, 9)</td>
<td>Tr(1, 1, 3)</td>
<td>Tr(3, 5, 7)</td>
</tr>
<tr>
<td>Control over manufacturing knowledge</td>
<td>Tr(2, 4, 6)</td>
<td>Tr(1, 1, 3)</td>
<td>Tr(8, 10, 10)</td>
</tr>
</tbody>
</table>

Note: For each input the first, second and third numbers represent the worst case, most likely case and best case scenarios respectively. Where only one number is present, that value remains constant through all scenarios.
Table 2.6. Assumptions for convertible debt financing.

<table>
<thead>
<tr>
<th>Term</th>
<th>10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>10 years</td>
</tr>
<tr>
<td>Coupon</td>
<td>5%</td>
</tr>
<tr>
<td>Conversion premium</td>
<td>20%</td>
</tr>
<tr>
<td>Probability of conversion before expiry date</td>
<td>0.5</td>
</tr>
<tr>
<td>Year of conversion</td>
<td>2, 7, 10</td>
</tr>
</tbody>
</table>

2.3.1. METHOD

The commercial and development characteristics of the drug candidates are outlined in Table 2.4. The input values for the 'Partner', 'CMO' and 'Build' options are detailed in Table 2.5. These values were determined based on assumed differences in capabilities between the options. For example, the CMO option was assumed to offer the highest titres due to their specialised expertise in the manufacture of biopharmaceuticals. Advice from industrial experts was solicited to ensure sensible assumptions were made. For the 'Partner/Build' option, manufacture of the first drug is modelled on the input values for the 'Partner' option and manufacture of the remaining drugs are modelled on the 'Build' option. Similarly, for the 'CMO/Build' option, manufacture of the first drug is modelled on the input values for the 'CMO' option and manufacture of the remaining drugs are modelled on the 'Build' option. The convertible debt instrument is the same across all scenarios and the input values can be found in Table 2.6. The expected values were used to generate a deterministic ranking configuration. By default the financial criteria and operational criteria were assigned equal aggregate weightings. The weighting of each group of financial criteria is equal. Likewise, the weighting of each group of operational criteria is equal.

The base case input values were processed through the decision-support framework to generate deterministic scores. A sensitivity analysis, based on the best and worst case values, was used to identify the most influential factors. A stochastic analysis was conducted using triangular distributions based on worst, most likely, and best values of manufacturing, commercial, and operational variables. Here, 1000 iterations of a Monte Carlo simulation were found to be sufficient to represent the stochastic behaviour of each option considered for the purposes of comparability.
2.4. RESULTS AND DISCUSSION

2.4.1. BASE CASE

Table 2.7. Deterministic results.

<table>
<thead>
<tr>
<th>Option</th>
<th>ENPV ($MM)</th>
<th>WSM Score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Build</td>
<td>$855</td>
<td>0.2157</td>
<td>1st</td>
</tr>
<tr>
<td>Partner</td>
<td>$641</td>
<td>0.2062</td>
<td>2nd</td>
</tr>
<tr>
<td>Partner/Build</td>
<td>$683</td>
<td>0.2026</td>
<td>3rd</td>
</tr>
<tr>
<td>CMO/Build</td>
<td>$829</td>
<td>0.1960</td>
<td>4th</td>
</tr>
<tr>
<td>CMO</td>
<td>$764</td>
<td>0.1795</td>
<td>5th</td>
</tr>
</tbody>
</table>

Table 2.8. Deterministic rankings of the options considered within each decision-making category.

<table>
<thead>
<tr>
<th>Group</th>
<th>Category</th>
<th>Partner</th>
<th>CMO</th>
<th>Build</th>
<th>Partner/Build</th>
<th>CMO/Build</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial</td>
<td>Profitability</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Asset Utilisation</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Liquidity</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Long-term Solvency</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Capital Structure</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Operational</td>
<td>Productivity</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Manufacturing Knowledge</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

The deterministic results are presented in Table 2.7, which for each option includes the WSM score, the assigned rank, and the ENPV generated. Table 2.8 highlights how the categories of criteria contribute to the score of each strategic option.

The Build, Partner and Partner/Build options scored higher than the average WSM score. The Build option scores highest indicating that is the most preferred option. This option generates the greatest ENPV value as well as the highest total value of assets which contribute highly in the capital structure category. Additionally, the highest degree of flexibility and the greatest expected gain in manufacturing knowledge are associated with this option.
The **Partner** option ranks second place. The sharing of expenses results in its total liabilities being the lowest of all the options and contributes to a high score in the long-term solvency category. The **Partner** option also involves sharing profits making it the least profitable option. The smaller profits result in a smaller equity value when compared with other options contributing to a low score in the capital structure category. This option also scored low in the flexibility category because the partnering company has been granted some control over the manufacturing projects.

The **Partner/Build** option scores lower than the **Partner** and **Build** options even though it is a combination of both. The benefits that the **Partner/Build** option gains from the **Build** option are counteracted because these benefits happen halfway into the lifetime of the project. This is relevant because the discount factor and the transition probabilities used to calculate the ENPV cashflows have the resultant effect of giving events occurring soonest the highest degree of impact on this figure. This will consequently impact the WSM score. Similarly, the fixed assets that are gained from building manufacturing capacity are used to generate highly discounted profits and hence lower equity. This results in a low score in the asset utilisation category and the capital structure category. Additionally, the **Partner/Build** option scores lowest in the productivity category as the labour force is not used as efficiently as with the **Partner** or **Build** options.

The **CMO** and **CMO/Build** options score the lowest overall. Contributors to these scores include the higher COGS and royalty charges associated with both of these options. This results in significantly higher current liabilities and lower equity values than any of the other options and gives rise to low scores in the liquidity and long-term solvency categories. Contracting out manufacturing obligations means that these options compromise some of the potential to build manufacturing knowledge so these options achieve lower scores in manufacturing knowledge. The **CMO/Build** option scores low in the productivity category as the labour force is not used as efficiently as with either the **CMO** or the **Build** options.
2.4.1. Sensitivity analysis

This analysis displays the maximal sensitivities across all options. For individual options these values will change as may the order of these factors. The analysis revealed that the most significant factor affecting the deterministic WSM score was the market capture achieved by the company (Figure 2.3). It was found that the most critical drivers were associated with commercial factors. These are similar to results from research conducted by Rajapakse et al. (2005) and Stonebraker (2002).

![Tornado diagram illustrating the overall maximal effects of best and worst case financial inputs on base case options.](image)

Figure 2.3. Tornado diagram illustrating the overall maximal effects of best (■) and worst (●) case financial inputs on base case options.

2.4.2. Expected Net Present Value (ENPV)

Use of the WSM proved to be highly useful in its ability to aggregate many aspects of the financial analysis and the operational evaluation. Furthermore, analysing the options under uncertainty via a Monte Carlo analysis also proved to be highly informative in revealing the expected performance of each option. Figure 2.4 demonstrates that the CMO/Build option is the most profitable and has a marginally greater upside potential in generating profit. The reason why the Build option does not generate the most profit in all circumstances is mainly due to its less developed technical manufacturing capabilities. In this case study the option to build has significantly less efficient technical manufacturing capabilities than with outsourcing.
to a contractor. This can create a greater barrier to generating profits in the *Build* option than the COGS premium and royalties charged will do in the *CMO/Build* option. When considering the *Build* and *CMO* options alone it can be seen that the *Build* option demonstrates a better performance in ENPV which is consistent with research by Ginsberg *et al* (2002). Additionally, it is shown that in this scenario increasing mean ENPV is coupled with greater uncertainty of this value being achieved.

![Distribution of the Expected Net Present Value](image)

**Figure 2.4.** Distribution of the Expected Net Present Value for the Partner (— ■ — ), CMO (•••••), Build (— ■ — ), Partner-Build (• • ) and CMO/Build (•••••••••••••) options under uncertainty.

### 2.4.3. Stochastic WSM scores

The stochastic WSM score highlights the holistic value of the option to the decision maker as defined by the decision-support methodology. Figure 2.5 reveals the difference in this case between an analysis based entirely on ENPV and one based on multiple criteria. In particular, the *CMO/Build* option which came first in the ENPV analysis is one of the lowest scoring options in the WSM analysis. Another notable difference is that none of the ranking positions achieved with the ENPV analysis are the same with the WSM analysis. In contrast, the stochastic WSM ranking positions are the same as the deterministic WSM ranking positions (Table 2.7 and Table 2.8).
Figure 2.5 also highlights a distinction between the top three options, Build, Partner, and Partner/Build and the remaining alternatives. By regarding the peaks of each distribution alone, the Build option can be considered to be most preferential and more certain, but not entirely distinct from the Partner option. The Partner/Build ranks closely behind the Partner and Build options but not close enough to be preferred over either. To draw further distinction between the Partner and Build options a two-sample two-tailed t-test assuming equal variances demonstrated full confidence that both samples are completely dissimilar. What needs to be determined from this point is the extent to which the WSM valuation can be relied upon to distinguish the Build and Partner options.

2.4.3.1. RISKS AND REWARDS

It is useful to quantify the impact of uncertain variables for each option on the WSM score. Here, the risk refers to the semi-standard deviation and the reward refers to the stochastic WSM score. Further, deviations above the mean value are not considered to be undesirable risk so the standard deviation which includes values both above and below the mean value, is not an appropriate metric. Figure 2.6 plots the semi-standard deviation versus the reward of each option and also displays the initial expenditure for each option. It is ideal to have options in the lower rightmost
quadrant. The Partner/Build option is shown to perform within this quadrant. The Build option performs approximately at the average semi-standard deviation value across the group with an above average reward.

![Diagram showing risk versus reward of considered options.](image)

**Figure 2.6.** Risk versus reward of considered options. The vertical dotted line and the horizontal dotted line are representative of the mean reward and risk values respectively across all options. The size of each bubble is proportional to the initial expenditure required to exercise each option.

The Build option is also shown to require the highest initial expenditure. If the company could not afford this option then it would have to reconcile between the tradeoffs seen with the other options that perform above the average reward. As demonstrated in Figure 2.6, these options would be the Partner and Partner/Build options. The Partner option offers a higher reward than the Partner/Build option but is also coupled with significantly higher risk.

This risk and reward analysis allows the decision maker to draw clearer distinctions between the Build and Partner options than was possible with Figure 2.4 or Figure 2.5 exclusively. Figure 2.6 makes clear that although the rewards are above average for the Build and Partner options, the Build option incurs a significantly lower risk. The Build option has the advantage of offering over 14% lower risk with nearly a 4% greater reward than the Partner option. It is intuitive to attach little advantage to this gain in reward however this is not necessarily an accurate association to make. As
seen in this case, a gain of nearly $200 million in ENPV is a contributor to this 4%
advantage in reward.

2.4.3.2. Financial and Operational Aggregate Scores

While analysing the holistic value of the WSM score reveals the overall worth of the
option to the decision maker it is important to discern the factors that have been most
influential in score’s determination. This is especially significant in confirming
whether or not the characteristics attributable to the highest ranking option are aligned
with the business strategy. Due to the structure of the MCDM based model it is
possible to breakdown the WSM value of each option into scores that relate
specifically to financial and to operational criteria. Figure 2.7 demonstrates the
balance of operational to financial scores for each option in this scenario. A high
financial score suggests a highly sustainable investment and likewise a high
operational score suggests a high level of practicality and manageability. Hence, the
most desirable option is the one that lies most to the upper rightmost quadrant. The
Build option performs fully above the group average for both scores. Comparison
between the Build and Partner options demonstrate that the Build option is most
valuable both financially and operationally.

Under uncertainty it is also highlighted that there is greater variance with the financial
aggregate scores than with the operational scores. Because of the subjective nature of
scoring for many of the operational attributes it would not be accurate to say that this
would necessarily translate to a lower operational risk.
Figure 2.7. Operational versus financial aggregate scores of the Partner (■), CMO (♦), Build (▲), Partner-Build (□) and CMO/Build (○). Each data point is representative of its deterministic value. The x and y error bars each signify one standard deviation either way of the data point. The vertical line intercepting the financial aggregate score axis is the mean financial aggregate score. The horizontal line intercepting the operational aggregate score axis is the mean operational score. The diagonal line shows the profile of financial and operational scores that are equally balanced.

2.4.3.3. OPERATIONAL TO FINANCIAL RATIO

The relative weighting of the operational aggregate score to financial aggregate score, $R$, is the final analysis considered in this chapter and is demonstrated in Figure 2.8. The previous analyses were under the presumption that operational and financial characteristics were of equal value to the decision maker. In Figure 2.8, values of $R$ ranging from 0 to 2 were used to investigate the stability of the ranking configuration. Within this range, it was found $R$ did not have an effect on the top ranking option demonstrating that this option dominates the other options when considering both the financial and operational aspects. It can be seen from Figure 2.7 that the 'CMO' option stochastically dominates all other options in operational scoring so as the value of $R$ increases this option becomes increasingly preferred and is most preferential at values of $R$ above 6.02.
Figure 2.8. The impact of the ratio, \( R \), of operational to financial aggregate weightings on the mean overall aggregate score for the Partner (■), CMO (♦), Build (▲), Partner-Build (□) and CMO/Build (○).

2.4.3.4. SUMMARY OF STOCHASTIC RESULTS

The results indicate that all options are financially and operationally viable but according to all analyses the Build option is most preferential. Although there were many conflicting attributes associated with the selection of any option, it is important to be aware that the Build option is the option most aligned with the company’s business strategy. Overall, the factors defining the decision proved to be its ability to exercise the option along with its willingness to accept the inherent tradeoffs. It is important to remember that the results are specific to the assumptions made in the case study; for example, if tighter budget constraints were assumed, this could influence the ranking of the option to build capacity and might even rule it out completely as infeasible.

Comparing the results presented, an investigation by Rajapakse et al. (2005) contained similar investigations with regard to risk and ENPV analysis. The paper involved the comparison of the construction of a biopharmaceutical manufacturing facility and the utilisation of a CMO. Similar to the results shown here was that the CMO displayed a significantly lower performance in ENPV when compared with building a new facility. Contrary to the results demonstrated here, was that the CMO
was found to be less risky than the option to build. It is difficult to compare results accurately as there are many differences between the case studies, assumptions and architectures of the models. For example, the drugs used in this case study are at later stages of development so the portfolio risk will be lower than a portfolio of drugs all entering early stage clinical trials as in Rajapakse et al. (2005). This reinforces the fact that the results are case-study specific.

One issue still to be resolved is that of the weightings of the criteria and their respective categories. In the case study equal weighting was assigned to all criteria and their categories but this configuration will conflict with the preferences of a decision maker that is more aggressive on generating profits. The line of inquiry following this is then in determining a suitable weighting configuration between the group, their categories, and their individual criteria to best represent the preferences of the decision maker. This consideration is outside the scope of this chapter.

Finally, the decision maker needs to be aware that, by definition, the normalisation of criteria values distorts or removes any meanings associated with their original magnitudes. Additionally, further distortion of the original magnitudes is exacerbated by the amalgamation of criteria values into a score and thus caution is required when interpreting the value of that score. As demonstrated earlier, the Build option outperformed the Partner option by a modest percentage in its WSM score but a significant gain in ENPV was a contributor in this performance.

2.5. CONCLUSIONS

The development of a decision-support framework to assist decision-making in strategies for the acquisition of biopharmaceutical manufacturing capacity has been presented. The framework provides a structured and transparent method of analysing such scenarios through the utilisation of MCDM and Monte Carlo simulation. Additionally, several financial and operational criteria were considered to provide a broad and detailed analysis from both perspectives. A hypothetical case study was formulated to demonstrate the usefulness and limitations of the framework.
The WSM proved to be highly suitable for data handling and for the analysis of results. The Monte Carlo simulation was valuable in highlighting the probability distributions and variance of base case values. Use of the model has highlighted that the employment of a single criterion in making strategic manufacturing decisions of this nature may not allow the decision maker to be aware of other important criteria. However, use of multiple criteria analysed under uncertainty provided a successful approach in identifying and confirming the best option. The analytical approach required highlights the complexity that can be involved in making decisions similar to the one analysed. Ultimately, a thorough and accurate analysis of financial and operational data is essential to make confident distinctions between feasible and attractive options.

### 2.6. NOMENCLATURE

- $\nu$: COGS at peak demand
- $a$: discount factor
- $b$: total sales
- $c$: coupon rate
- $C$: criterion
- $e$: attribute value
- $E$: total annual expense
- $E_1$: capital investment
- $E_2$: cost of goods sold
- $E_3$: general and administrative
- $E_4$: royalties
- $E_5$: total plant expenses to be paid
- $F$: fixed capital investment
- $g$: proportion of sales to be spent on general and administrative costs
- $I$: total annual income
- $I_1$: dose per patient per year
- $I_2$: target population size
- $I_3$: market capture
- $I_4$: annual product demand
A MULTI-CRITERIA DECISION-MAKING FRAMEWORK FOR THE SELECTION OF STRATEGIES FOR
ACQUIRING BIOPHARMACEUTICAL MANUFACTURING CAPACITY

$I_5$ estimated sales price
$I_6$ other income
$L$ par value of issued convertible debt
$m$ proportion of profit payable as tax
$M$ peak market capture
$N$ total number of drugs
$p$ proportion of profits owed to partner
$P_I$ profit before interest and tax
$P_2$ depreciation
$P_3$ taxable profit
$P_4$ tax
$p_e$ proportion of expenses that the company is obligated to pay
$P_{ENPV}$ expected net present value
$P_{NPV}$ net present value
$P_{PV}$ annual present value
$P$ profit after interest and tax
$q$ number of manufacturing personnel required
$r$ proportion of sales that will used to pay royalty fees
$t_D$ decay period after peak market capture
$t_R$ ramp period from commercialisation to peak market capture
$w$ normalised weight value
$x$ normalised attribute value
$Y$ project duration in years
$z$ weight value
$a$ total assets
$\alpha_C$ current assets
$\lambda$ total liabilities
$\lambda_C$ current liabilities
$\pi_c$ cumulative transition probability
$\pi$ transition probability
$\chi_{F,k}$ $k^{th}$ qualitative component of the flexibility category
$\chi_{L,k}$ $k^{th}$ qualitative component of the location category
$\chi_{MA,k}$ $k^{th}$ qualitative component of the manufacturing knowledge category
$\chi_{P,k}$ $k^{th}$ qualitative component of the productivity category
Subscripts:

\( j \) \( j^{th} \) alternative option

\( i \) \( i^{th} \) criterion

\( t \) year \( t \)

\( n \) drug \( n \)
CHAPTER 3

STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT

3.1. INTRODUCTION

In chapter 2 a framework for reconciling between strategic options for acquiring biomanufacturing capacity was presented. The work in this chapter takes this considerably further by exploring the options available to a biopharmaceutical manufacturer for the development of a portfolio of drugs and the development of a technique to optimise such strategies. Given the relevance and importance of contributions to large-scale portfolio and capacity optimisation problems in biopharmaceutical settings, especially in modern industry, this chapter addresses the development of a holistic framework for the simultaneous stochastic combinatorial and multi-objective optimisation of biopharmaceutical research and development portfolio management and manufacturing capacity planning decisions. More specifically, given a set of drug candidates and their uncertain development, manufacturing, and commercial parameters, alongside an availability of external corporate bodies for development and manufacturing with their various uncertain technical characteristics this work presents a novel method for finding the optimal structure of the:

- Optimal structure of the drug development portfolio,
- Development sequence for the selected drug candidates,
- Schedule of critical development activities, and
- Itinerary of activities at specific stages that should be integrated in-house, outsourced, or partnered such as to maximise important multiple objectives.

The overall framework is a combination of the simulation-based evaluation framework based on the framework described in chapter 2 and a bespoke estimation of distribution algorithm (EDA) (Mühlenbein and Pass, 1996) to iteratively evolve a population of candidate strategies. All strategies are formulated according to a common superstructure and the simulation model, which integrates a
biopharmaceutical manufacturing process model, is used to determine important stochastic properties demonstrated in performance objectives. Due to the vast multidimensional decision space and multimodal objective space presented by the addressed problem, this work features an EDA that harnesses the unsupervised machine learning capabilities of Bayesian networks to discover the probabilistic nature and interconnected structure of the decisions that comprise superior performing strategies within the model. To further support the capture of probabilistic data and hence the representational power of the EDA, superior strategies are algorithmically clustered according to their performance in the objective space so that a specific probabilistic model can be constructed for each clustered region. The EDA then uses these probabilistic models to formulate new strategies. Each iteration in the optimisation procedure consists of evaluating a population of candidate strategies, selection of superior candidate strategies, probabilistic modelling and generation of the subsequent population. The approach also features a range of uncertain parameters, product line and corporate affiliation dependency specifications. In addition to these structural elements this study pays attention to the practicalities of illustrative cases. It is the aim of this chapter to investigate how performance optimised strategies and the boundaries of optimal performance in the objective space will vary according to selected constraints. These constraints are the size of the portfolio and the magnitude of the constraint on cash flow.

The organisation of the remainder of this chapter is now described. Section 3.2 details the complete model formulation for the stochastic evaluative framework and for the EDA. Section 3.3 presents a case specific to the production of monoclonal antibody drug candidates to exemplify the use of the framework. Section 3.4 describes and discusses the results pertaining to the case presented. The final section discusses any final conclusions.

3.2. MODEL FORMULATION

The stochastic optimisation process makes use of an evaluative framework to capture the impact of decisions made by strategies in the modelled environment and to estimate their stochastic properties. This is coupled with an EDA that operates the optimisation procedure through the machine learning of instances of decision
variables that are associated with superior performance. Here, any interdependent relationships between decision variables and their instances are data mined and then used to build a probabilistic model on which the formulation of new strategies are based.

The model is designed using C++ and MS Excel. The entire simulated environment is assembled in C++ as this contains the most frequently utilised calculations and the C++ language offers significant savings in computational time. MS Excel is used as the main graphical user interface as it offers a convenient means for storing, extracting and manipulating data. The environment is interfaced with MS Excel via a dynamic link library where it is represented as a user-defined function. Decision variables are entered into the function and defined parameters are outputted. Modules for controlling the flow of data in the simulation and optimisation procedures are compiled in Visual Basic for Applications (VBA) which is readily executable in the MS Excel environment.

To illustrate to the reader how the salient sub-procedures of the framework link together, an example describing the flow of computational tasks is henceforth summarised. This should also provide the reader with an overall perspective of the methodology and an understanding of how the following subsections are linked. A batch of strategies is generated towards a simulated environment that characterises the portfolio problem that is to be investigated. Each strategy in this batch is sequentially selected and simulated where the decisions for portfolio construction, scheduling, and the involvement of third parties are recognised by the framework. Simulation tasks include calculating the economics of development, manufacturing, and marketing which will lead to the determination of objective metrics. The simulation framework contains stochastic variables which require the simulation to be repeated for multiple trials for each strategy. This will subsequently provide statistics pertaining to the objective metrics that are useful for comparing the performance of different strategies. Once all strategies have been simulated an algorithmic selection process is used to identify superior strategies according to their performance against pre-determined objectives whilst discarding the remainder. These superior strategies are split into a specified number of clusters, again based on their performance against pre-determined objectives. For each of these clusters a Bayesian network is learned in order to
identify and classify any probabilistic interdependencies between the decisions that comprise the member strategies. This provides an understanding, albeit complex and from a probabilistic perspective, as to how strategies in each cluster and hence superior strategies are structured. The learned Bayesian networks serve as probabilistic models which are sampled in order to generate new and improved strategies. Here the entire batch of strategies that was previously simulated is now replaced with a new batch that is modelled on its superior features. This simulation and regeneration procedure is repeated until a termination criterion is reached. Thorough details of the entire method are now described.

3.2.1. EVALUATIVE FRAMEWORK

The framework for evaluating each candidate solution (Figure 3.1) models the performance of the strategy by mapping the consequences of decisions made within the simulated environment. Various facets of the drug development process and its wider commercial environment are captured within the modular structure of the evaluative framework. Each module is designed to carry out a set of related calculations and, as illustrated, these modules are networked and eventually result in
the evaluation of a set of criteria. The following sub-sections will detail the contents of these modules.

### 3.2.1.1. Super-structure of a Candidate Solution

Consider generation $G(t)$ as a population of candidate solutions at a given iteration of the optimisation procedure, $t$, within the total number of iterations, $t_{MAX}$, that can potentially maximise the objectives of the decision maker. Each candidate solution, $g$, in $G(t)$ has a generalised structure for representing its decisions. Figure 3.2 demonstrates this generalised structure for a portfolio of five drugs where the structure can be segregated into three sub-types: drug development sequence, scheduling, and third party usage.

The strategy for the drug development sequence, $D_g$, codes the drugs that are included in the portfolio and the order in which they are commercialised:

$$D_g = \{D_{g,i=1}, D_{g,i=2}, D_{g,i=3}, ..., D_{g,i=I}\} \quad \forall g \in G(t), i \in I \quad (3.54)$$

where $D_{g,i}$ is the drug chosen as the $i^{th}$ drug in the drug development sequence by the $g^{th}$ candidate solution, and $I$ is the total number of drugs in the drug development portfolio. It is clear that $i$ can be any drug not already chosen as part of $D_g$.

The scheduling strategy for the $i^{th}$ drug in the development sequence according to the $g^{th}$ candidate solution, $T_{g,i}$, commences development of drug $i$ according to when drug $i-1$ reaches the beginning of a particular stage of development. These stages can be instantiated as: target identification ('ID'), preclinical testing ('PC'), phase I clinical trials ('PI'), phase II clinical trials ('PII'), phase III clinical trials ('PIII'), FDA review ('FDA'), and market approval ('MKT'). Thus:

$$T_g = \{T_{g,i-1}, T_{g,i-2}, T_{g,i-3}, ..., T_{g,i-1}\} \quad \forall g \in G(t), i \in I \quad (55)$$

$$T_{g,i} = \{'ID', 'PC', 'PI', 'PII', 'PIII', 'FDA', 'MKT'\} \quad \forall g \in G(t), i \in I \quad (56)$$

where $T_g$ is the set of timing strategies for all drugs in the portfolio.
The third party strategy belonging to the \( g \)th solution for the \( i \)th drug at the \( j \)th development activity, \( C_{g,i,j} \), has three possible instances for each stage: in-house or otherwise referred to as an integrated activity ('I'), outsourced activity ('C'), and partnered activity ('P'), hence:

\[
C_g = \{C_{g,i-1}, C_{g,i-2}, C_{g,i-3}, \ldots, C_{g,i}\} \quad \forall g \in G(t), i \in I \tag{3.57}
\]

\[
C_i = \{C_{g,i-1}, C_{g,i-2}, C_{g,i-3}, \ldots, C_{g,i}\} \quad \forall g \in G(t), i \in I, j \in J \tag{3.58}
\]

\[
C_{g,i,j} = \{ 'I', 'C', 'P' \} \quad \forall g \in G(t), i \in I, j \in J \tag{3.59}
\]

where \( J \) is the number of development activities. This format allows the framework to formulate a plan of critical activities to complete either by themselves or by another corporate body. The development activities, \( C_{g,i,j} \), are structured as: \( C_{g,i,1} \) - target identification, \( C_{g,i,2} \) - preclinical testing, \( C_{g,i,3} \) - phase I clinical testing, \( C_{g,i,4} \) - phase II clinical testing, \( C_{g,i,5} \) - phase III clinical testing, \( C_{g,i,6} \) - manufacturing for phase I clinical trials, \( C_{g,i,7} \) - manufacturing for phase II clinical trials, \( C_{g,i,8} \) - manufacturing for phase III clinical trials, and \( C_{g,i,9} \) - commercial manufacturing.

Figure 3.2. The super-structure of a candidate strategy for the commercialisation of a portfolio of five drugs.
Structuring the strategies $D_g$ and $T_g$ is straightforward. Each instance within $D_g$ can only refer to one drug whilst any combination of timing instances for each $T_{g,i}$ in $T_g$ is permissible. Structuring the corporate strategy requires the application of some rules to avoid the formulation of nonsensical strategies. The first set of stipulations restricts the company from breaking and resuming contracts with partners:

If $C_{g,i,1} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{2, 3, 4, 5, 6, 7, 8, 9\}$ (3.60)

If $C_{g,i,2} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{3, 4, 5, 6, 7, 8, 9\}$ (3.61)

If $C_{g,i,3} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{4, 5, 6, 7, 8, 9\}$ (3.62)

If $C_{g,i,4} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{5, 7, 8, 9\}$ (3.63)

If $C_{g,i,5} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{8, 9\}$ (3.64)

If $C_{g,i,6} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{3, 4, 5, 7, 8, 9\}$ (3.65)

If $C_{g,i,7} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{4, 5, 8, 9\}$ (3.66)

If $C_{g,i,8} = 'P'$ then $C_{g,ij} = 'P'$ for $j = \{5, 9\}$ (3.67)

This assumes that the partner is required for both clinical trial development and manufacturing. A second set of stipulations prevents the breaking of outsourcing contracts on clinical trial and clinical manufacturing activities:

If $C_{g,i,3} = 'C'$ then $C_{g,ij} = 'C'$ for $j = \{4, 5\}$ (3.68)

If $C_{g,i,4} = 'C'$ then $C_{g,ij} = 'C'$ for $j = \{5\}$ (3.69)

If $C_{g,i,6} = 'C'$ then $C_{g,ij} = 'C'$ for $j = \{7, 8\}$ (3.70)

If $C_{g,i,7} = 'C'$ then $C_{g,ij} = 'C'$ for $j = \{8\}$ (3.71)

It is assumed that contracting for clinical trial testing and contracting for manufacturing will be completed by separate companies.

These stipulations have an added benefit of reducing the decision space and computational time. Without them each drug would have a total decision space for $C_{g,i}$ of $3^9$ solutions over its nine constituent decision variables, now this is reduced to 207 solutions. To support computational efficiency $C_{g,i}$ is programmed as a single variable having 207 instances, rather than as nine $C_{g,ij}$ variables each having three
instances. This format means that $C_{g,i}$ can be instantiated from a database of permissible combinations of each $C_{g,j,i}$, as opposed to having an active procedure that validates and corrects the structure of $C_{g,i}$ for each $g$ in $G(t)$ over each $t$ through to $t_{\text{MAX}}$.

3.2.1.2. STOCHASTIC VARIABLES

To simulate the uncertain environment the model accounts for the stochastic nature of a range of uncertain variables (Table 3.1). In this study, each stochastic variable is characterised by a triangular probability distribution based on the specification of maximum, minimum, and most likely values. The types of variables included in the model formulation are costs, commercial factors, and manufacturing capabilities. Each variable bears relevance to a single element of $C_{g,i}$ and a separate distribution must be defined for each possible instance of the relevant element. Thus, for a portfolio of five drugs, individual distributions of 99 stochastic variables must be defined. Depending on the structure of $g$ the relevant stochastic variables will be selected for sampling.
Table 3.1. Stochastic variables.

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Variable</th>
<th>Relevant Sub-Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target identification</td>
<td>$C_{g,i,1}$</td>
</tr>
<tr>
<td></td>
<td>Preclinical testing</td>
<td>$C_{g,i,2}$</td>
</tr>
<tr>
<td></td>
<td>Phase I clinical trials</td>
<td>$C_{g,i,3}$</td>
</tr>
<tr>
<td></td>
<td>Phase II clinical trials</td>
<td>$C_{g,i,4}$</td>
</tr>
<tr>
<td></td>
<td>Phase III clinical trials</td>
<td>$C_{g,i,5}$</td>
</tr>
<tr>
<td></td>
<td>Manufacturing for phase I clinical trials</td>
<td>$C_{g,i,6}$</td>
</tr>
<tr>
<td></td>
<td>Manufacturing for phase II clinical trials</td>
<td>$C_{g,i,7}$</td>
</tr>
<tr>
<td></td>
<td>Manufacturing for phase III clinical trials</td>
<td>$C_{g,i,8}$</td>
</tr>
<tr>
<td></td>
<td>Commercial manufacturing</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Scale-up synthesis</td>
<td>$C_{g,i,6}$</td>
</tr>
<tr>
<td></td>
<td>Formulation</td>
<td>$C_{g,i,7}$</td>
</tr>
<tr>
<td></td>
<td>Commercial Preparation</td>
<td>$C_{g,i,8}$</td>
</tr>
<tr>
<td></td>
<td>Marketing</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>FDA Review</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td>Cost</td>
<td>Demand</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td></td>
<td>Time required to ramp to peak market</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Time taken for full market decay</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Compound annual growth rate</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td></td>
<td>Sales price uncertainty</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td>Commercial Factors</td>
<td>Demand</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td></td>
<td>Time required to ramp to peak market</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Time taken for full market decay</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Compound annual growth rate</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td></td>
<td>Sales price uncertainty</td>
<td>$D_{g,i}$</td>
</tr>
<tr>
<td>Manufacturing Capabilities: Phase I Clinical Trials</td>
<td>Whole process yield</td>
<td>$C_{g,i,6}$</td>
</tr>
<tr>
<td></td>
<td>Fermentation titre</td>
<td>$C_{g,i,6}$</td>
</tr>
<tr>
<td></td>
<td>Probability of a successful batch</td>
<td>$C_{g,i,6}$</td>
</tr>
<tr>
<td>Manufacturing Capabilities: Phase II Clinical Trials</td>
<td>Whole process yield</td>
<td>$C_{g,i,7}$</td>
</tr>
<tr>
<td></td>
<td>Fermentation titre</td>
<td>$C_{g,i,7}$</td>
</tr>
<tr>
<td></td>
<td>Probability of a successful batch</td>
<td>$C_{g,i,7}$</td>
</tr>
<tr>
<td>Manufacturing Capabilities: Phase III Clinical Trials</td>
<td>Whole process yield</td>
<td>$C_{g,i,8}$</td>
</tr>
<tr>
<td></td>
<td>Fermentation titre</td>
<td>$C_{g,i,8}$</td>
</tr>
<tr>
<td></td>
<td>Probability of a successful batch</td>
<td>$C_{g,i,8}$</td>
</tr>
<tr>
<td>Manufacturing Capabilities: Market</td>
<td>Whole process yield</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Fermentation titre</td>
<td>$C_{g,i,9}$</td>
</tr>
<tr>
<td></td>
<td>Probability of a successful batch</td>
<td>$C_{g,i,9}$</td>
</tr>
</tbody>
</table>
3.2.1.3. Biomanufacturing model

The framework used to model the manufacture of biopharmaceutical products is based on work reported by George et al. (2007), Farid et al. (2000, 2005a, 2006), and Lim et al. (2005a and b). Included in the model are the main process and ancillary tasks involved in the manufacturing procedure. The main input parameters to the model are the annual demand to be met by the facility, the expected fermentation titre, the overall product yield, and the probability of achieving success for a single batch fermentation. Equations for calculating the utilisation of equipment, materials, utilities and labour are also included. Cost calculations are supported by a database of unit costs for equipment and materials used in the manufacturing process.

3.2.1.4. Dependencies

Within a set of decisions it is possible that at least one of these decisions can have an impact on the performance of remaining decisions that are yet to be executed. In the real world, this can be a consequence of making decisions that affect the utilisation of the same tangible or intangible resource. The framework recognises three such contexts of dependency where this type of impact may occur: contractual, revenue, and manufacturing cost.

Contractual dependencies here refer to the utilisation of a third party for manufacturing or research activities. Here, the premise is that the longer the period for which a third party is used and the greater the number of activities that it is involved in, the more favourable the rates it charges becomes. Additionally, for each third party there is a minimum charge that it will not breach. This clearly affects how the corporate relation strategy performs in the model.

Revenue dependencies here refer to the impact that constituents of the company’s drug pipeline create when competing within the same market. When multiple drugs compete in the same market each drug can suffer from reduced returns in comparison to what it might have achieved if commercialised in absence of competing drugs, given the same commercial environment. It is also possible that multiple drugs can
enhance the sales revenue of each other if they have complementary applications. In
the model, the impact of this dependency is realised through \( D_g \) and \( T_g \). The
configuration of \( T_g \) is important here because, in the absence of other dependencies
affected by these strategies, it may be more beneficial to stagger the development of
competing drugs so that there are periods of reduced or no competition. Also, the
performance of these strategies will depend on the revenue related penalties or
benefits of having competing or complementary drugs in the market place.

Capital dependencies are modelled here as affecting the capital expense required for
manufacturing a drug, due to the sharing of resources for structurally similar drugs.
Where such drugs are being manufactured, it may be possible to use some part of the
same manufacturing facilities hence reducing the overall capital requirements. The
impact of this dependency is reliant on the capital savings than can be realised within
a group of drugs. This dependency is affected by the strategies for the structure of the
drug portfolio where the choice of structurally similar drugs supports reduced capital
expense requirements. The strategy for the order in which drugs are commercialised
is also important because drugs commercialised later in the pipeline may require less
capital. Additionally, these strategies must reconcile the penalties and benefits of
their impact on capital and revenue dependencies.

### 3.2.1.5. Processing selection criteria and stochastic properties

The simulation of any candidate strategy will lead to a distribution of outcomes in
each performance criterion considered, providing a variety of information to the
decision maker. As a basis for comparing the quality and superiority of strategic
options the following approach was taken. The approach was based on two metrics,
the mean positive NPV produced by the strategy and the probability of generating a
NPV above zero, \( p(\text{NPV}>0) \). These allow for the decision maker to establish which
strategies demonstrate optimal performance in generating profit and those that
maximise the expectancy of generating profit of any magnitude. This was also used
for the selection of superior strategies for which the procedure will be formalised in
later sections.
The rationale for bifurcating the distribution of NPV values is now discussed. The distribution of all NPV values produced by each strategy is expected to be bimodal. The leftmost mode will be correspondent to the expected level of loss or the expected impact of failure when the candidate strategy does not yield a positive NPV. Similarly, the rightmost mode will correspond to the expected level of financial gain when the strategy yields drug development projects that successfully reach the marketing stage. It should be borne in mind that the rightmost mode is expected to be positive for all strategies, and this may especially be true in scenarios where heavy development costs and particularly low probabilities of reaching the marketing stage are combined with imprudent third party strategies and projects that yield modest revenues. Additionally, as realistic expectancies of a monoclonal antibody drug progressing to the marketing phase are typically below 0.3 (Reichert, 2003) the leftmost mode is expected to be the most pronounced of the two. In the optimisation framework, if the NPV distribution is treated as a single attribute then the optimisation process is more likely to be driven towards minimising the impact of failure than it is towards maximising the rewards that come with successful drug development. Hence, the NPV distribution is treated as two separate attributes of the candidate strategy that are simply dichotomised at NPV = 0. If an alternative objective function was used then it may be appropriate to consider the entire distribution rather than any specific and separable constituents of it. Additionally, it should be considered that the probability of achieving a positive NPV is a single metric derived from the simulation data and clearly does not have a mean, minimum, maximum or semi standard deviation.

The objective function $p(NPV>0)$ is a straightforward calculation that can also be extended to the more specific needs of the decision maker. For example $p(NPV>i)$ or $p(NPV>j) > j$ could be used as suitable and more malleable alternatives, where $i$ and $j$ used in this context are defined by the needs of the decision maker. The objective function $p(NPV>0)$ is used here so that the full landscape of the objective space can be viewed and analysed.
3.2.2. Estimation of Distribution Algorithm

EDAs, otherwise known as estimation of distribution algorithms (EDAs) (Mühlenbein and Paß, 1996), are comparable to a class of optimisation algorithms known as genetic algorithms (GAs) (Holland, 1975) and have been shown to be a class of promising approaches in solving combinatorial optimisation problems (Pelikan et al., 2000). Like GAs, EDAs function by iteratively evolving a population of candidate solutions to the problem until a termination criterion is satisfied. A main driver by which GAs are thought to achieve this is through the manipulation of building blocks (Holland, 1975; Goldberg, 1989). Building blocks are a central concept in GA theory, where superior building blocks are expected to be key components of superior solutions. They can be any set of instantiations of decision variables present in any set of candidate solutions. Additionally, a superior building block can be considered as a set of instances of decision variables that work together to support the performance of any set of superior candidate solutions in the objective space considered. The instances of decision variables forming a building block may also exhibit a dependency relationship where added value in superiority is achieved by the holistic presence of a set of instances rather than necessarily through individual incremental increases in performance contributed by each constituent of the building block. EDAs are designed to recognise superior building blocks explicitly by constructing probabilistic models of the states of decision variables that are present in superior solutions. Examples of simple EDAs in literature include the univariate marginal distribution algorithm (UMDA) (Mühlenbein, 1997) and the compact genetic algorithm (cGA) (Harik et al., 1999). Examples of more complex EDAs include the Bayesian optimisation algorithm (BOA) (Pelikan et al., 2000) and the multi-objective hierarchical Bayesian optimisation algorithm (mohBOA) (Pelikan et al., 2005).

In this study, and like the BOA, the data structure used as the framework for the probabilistic model is a Bayesian network. A Bayesian network is an annotated directed graph that encodes probabilistic relationships and their use here derives from the fact that artificial intelligence researchers have used them to encode expert knowledge (Heckerman, Geiger and Chickering, 2005). The topology of the network is learned directly from the structure of top performing candidate solutions in $S(t)$ and
is then randomly sampled via the conditional probabilities that it encodes to generate 
$G(t+1)$. The EDA used here (Figure 3.3) is largely based algorithmic concepts found 
in the mohBOA.

**Estimation of Distribution Algorithm**

*Initial population:*

- Let the population of candidate solutions be $G(t)$.
- $t = 1$
- Randomly generate the initial population of candidate solutions, $G(1)$.
- Let $t_{\text{max}}$ be the maximum number of populations to be evolved.

For $t = 1$ to $t_{\text{max}}$:

**Evaluation:**

- Simulate each $g$ in $G(t)$ and record stochastic properties.

**Selection:**

- Use the fast nondominated sorting and crowding distance algorithms to select the top 50% 
of solutions, $S(t)$, from $G(t)$.

**Clustering of the objective space:**

- Using the k-means clustering algorithm, separate $S(t)$ into $z$ clusters.

For each cluster, $K_z$:

**Probabilistic model building:**

- Construct the Bayesian network, $B_z$, using a hill climbing procedure to optimise the 
  Bayesian Dirichlet metric over $B_z$.

**Sampling of the probabilistic model:**

- Generate a new set of strategies $O_z(t)$ by randomly sampling the joint probability 
  distribution encoded by $B_z$. The number of strategies to be generated will be twice the 
  original cluster size.

**Population regeneration:**

- Generate the new population $G(t+1)$ by randomly by replacing all strategies in $G(t)$ with all strategies 
in each $O_z(t)$.

**Figure 3.3.** Pseudocode for the EDA.
3.2.2.1. Fast non-dominated sorting

Fast nondominated sorting (Deb et al., 2002) is used to support the discovery of a wide spread and densely populated Pareto optimal front (Figure 3.4). This is of great value to the optimisation process where discovery of a wide spread Pareto front at each iteration can reduce the likelihood that the optimisation algorithm converges to a small and restricted region of the objective space and also increases the capacity for the algorithm to discover new areas of the objective space. Algorithmically steering the discovery towards candidate solutions along the Pareto front has its advantages. Firstly, the nature and severity of conflict between objectives can be better understood with wider coverage of the objective space. Secondly, a more efficient traversing of the objective space can be achieved through accelerated progress at individual iterations. Thirdly, improved probabilistic models are likely to be built as they are more representative of candidate solutions along the Pareto front. This also supports the likelihood that stronger performing candidate solutions will be sampled from these improved models. Nondominated sorting is designed to support achievement of these advantages by splitting the objective space into Pareto layers, and then identifying the members of each layer. Pareto layers are comprised of candidate solutions that are Pareto optimal in the absence of other dominating solutions. A candidate solution is Pareto optimal if there is no other solution that exists which yields an improvement in any objective without yielding degradation in any other objective. The first layer contains candidate solutions that are not dominated by any other candidate solution in \( G(t) \). In the absence of members belonging the first layer, members of the second layer will not be dominated by any of the remaining solutions. Accordingly, the layer to which each member belongs represents its rank so those in the first layer will be the most highly ranked in \( G(t) \).
Fast Nondominated Sorting Algorithm

For each candidate solution \( p \) in the current population \( G(t) \):

- Empty the set of solutions that are dominated by \( g \), \( S_g \)
- The number of solutions that dominate \( g \), \( n_g = 0 \)

For each remaining solution \( q \) in \( G(t) \):

- If \( g \) dominates \( q \) then:
  \[ S_g = S_g \cup \{ q \} \]
- Else if \( q \) dominates \( g \) then:
  Increment the domination count of \( g \), \( d_g = d_g + 1 \)

If \( d_g = 0 \) (\( g \) is a nondominated solution) then:
- The domination rank of \( g \), \( r_{\text{rank}} = 1 \)
- Add \( g \) to the set forming the first Pareto layer, \( F_1 \)

\( f = 1 \)

While the set of members for the \( f \)th Pareto layer, \( F_f \), is not empty:

- Set \( Q \) as having no members.
- For each \( g \in F_f \):
  - For each \( q \in S_g \):
    - Decrement the domination count of \( g \), \( d_g = d_g - 1 \)
    - If \( d_g = 0 \) then:
      - The domination rank of \( q \), \( r_{\text{rank}} = f + 1 \)
- \( Q = Q \cup \{ q \} \)
- \( f = f + 1 \)
- \( F_f = Q \)

Figure 3.4. Pseudocode for the nondominated sorting procedure. Adapted from Deb et al. (2002).

3.2.2.2. CROWDING DISTANCE

To support the function of the nondominated sorting procedure, crowding distance is used to determine how densely populated the current objective space around this solution is (Figure 3.5). (Deb et al., 2002). The principle is that candidate solutions situated in less densely populated areas of the objective space are of higher preference than those in more densely populated areas. The reason being is that a candidate solution in a less densely populated region is considered to be more unique. Also, the more unique candidate solutions that are included when building probabilistic models the more diverse the base of information will be when characterising the structure of superior candidates. Thus, a candidate solution with a relatively high crowding
distance will be situated in a region of the objective space that is relatively less dense and will be considered to be relatively more unique. The measure of crowding distance for a candidate solution is the perimeter of the cuboid created by its two most proximal candidate solutions in the objective space.

**Crowding Distance Algorithm**

Let $|G(t)|$ be the number of candidate solutions in $G(t)$.

For each candidate solution $g \in G(t)$:

Crowding distance for $g$, $C(g) = 0$

For each objective, $v$:

Let $\Phi$ be the set of candidate solutions in $G(t)$ sorted in ascending order according to $v$.

$C(\Phi[1]) = \infty$

$C(\Phi[|G(t)|]) = \infty$

Let $\mu(g)$ be the value of the $v^{th}$ objective as a function of $g$.

For $g = 2$ to $(|G(t)| - 1)$:

$C(\Phi[g]) = C(\Phi[g]) + (\mu(\Phi[g + 1]) - \mu(\Phi[g - 1]))(\mu(\Phi[1]) - \mu(\Phi[|G(t)|]))^1$

**Figure 3.5.** Pseudocode for the allocation of crowding distances. Source: Adapted from Deb et al. (2002).

Both the fast nondominated sorting and crowding distance procedures form the basis for the selection of successful candidate solutions on which the generation of $G(t+1)$ will be based. Figure 3.6 describes the procedure by which this is achieved. Fast nondominated sorting is used to select candidate solutions in nondominated layers and add them to the set of superior performing candidate solutions, $S(t)$. If all the members of the next Pareto layer will not fit into $S(t)$ then candidate solutions are chosen in order of having the greatest crowding distance.
STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT

Selection of Superior Candidate Solutions

Let \( S(t) \) be the set of successful candidate solutions to be selected from the current population, \( G(t) \).

Let \( |S(t)| \) be the number of successful candidate solutions to be selected.

Let \( ||S(t)|| \) be the number of candidate solutions remaining to be selected.

\[ ||S(t)|| = |S(t)| \]

Sort all members of \( G(t) \) using the fast nondominated sorting procedure.

For each candidate solution, \( g \) in \( G(t) \), determine its crowding distance, \( C(g) \).

Let \( F^f \) be the set of candidate solutions belonging to the \( f \)th Pareto layer.

For each \( f \)

Let \( |F_f| \) be the number of candidate solutions in \( F_f \).

If \( |F_f| \leq ||S(t)|| \) then:

\( S(t) = S(t) \cup \{ F_f \} \)

\[ ||S(t)|| = ||S(t)|| - |F_f| \]

Else:

Sort all members of \( F_f(t) \) according to crowding distance.

Let \( G(t)_{\text{select}} \) be the set of \( ||S(t)|| \) candidate solutions belonging to \( G(t) \) with the greatest crowding distance values.

\[ S(t) = S(t) \cup \{ G(t)_{\text{select}} \} \]

Figure 3.6. Pseudocode for the selection of superior candidate solutions.

3.2.2.3. K-MEANS CLUSTERING

Once \( S(t) \) has been populated, it is algorithmically clustered so that the properties of superior candidate solutions can be analysed for region-specific properties in the objective space (Figure 3.7). The k-means clustering algorithm (MacQueen, 1967) was selected for this purpose because its ease of implementation. The advantage of clustering the objective space is that it supports the avoidance of identifying traits of candidate solutions in one region of the objective space that may not support the performance of traits present in other regions. This algorithm requires the number of clusters and an initial set of coordinates that represent the centres of these clusters in the objective space to be specified by the decision maker. The initial coordinates can be set randomly. Members of \( S(t) \) are each assigned to a cluster based on the corresponding cluster centre of the closest Euclidean distance. The coordinates for each cluster centre is sequentially updated until the set of members for each cluster centre is unchanged when compared to the previous iteration.
**k-means Clustering Algorithm**

Input the number of clusters, $z$.

Initialise a coordinate for each cluster centre, $k$, where each dimension in $k$, $k_{z,v}$, is a value for each performance objective, $v$.

Let $K_{z,h}$ be the set of candidate solutions in the $z^{th}$ cluster at the $h^{th}$ iteration.

$h = 0$

Do until $K_{1,h} = K_{1,h-1}, K_{2,h} = K_{2,h-1}, ..., K_{z,h} = K_{z,h-1}$:

$h = h + 1$

For each candidate solution, $g$, in the current population, $G(t)$:

Initialise each $K_{z,h}$ as having no members.

Let $\mu_v(g)$ be the value of the $v^{th}$ objective as a function of $g$.

For each $z$:

Determine the Euclidean distance of $p$ from $k$, $e_{z,v}$:

$$e_{z,v} = \sqrt{\sum_{v} \left( \mu_v(g) - k_{z,v} \right)^2}$$

For the $k_z$ that produces the smallest $e_{z,v}$:

$K_{z,h} = K_{z,h} \cup \{g\}$

For each $z$:

For each $v$:

Let $|K_{z,h}|$ be the number of candidate solutions in $K_{z,h}$.

$$k_{z,v} = \left| K_{z,h} \right| \sum_{n=1}^{\left| K_{z,h} \right|} \mu_v(g_n)$$

Figure 3.7. Pseudocode of the k-means clustering algorithm.

### 3.2.2.4. Constructing the Bayesian Network

Bayesian networks (Neapolitan, 2003) are used here to characterise the probabilistic nature of relationships between decision variables and their instances occurring in the set of candidate solutions of the $z^{th}$ cluster, $K_z$, via machine learning. This data structure is subsequently used for generation of $G(t+1)$. As the objective space is partitioned into $z$ clusters a Bayesian network, $B_z$, is constructed for each $z$. Each decision variable, $\delta_n$, is represented by a node, $\delta_{nz}$, in the network $B_z$. Any relationships between each $\delta_{nz}$ are represented by a directed edge drawn between them (see Figure 3.8 as an example). The node from where the edge is directed is a parent node, while the node to which the edge is directed is its child node. That is, within each $B_z$ the instance of each child node is conditional upon the set of instances of its parents. It is important to note that all $B_z$ are acyclic so a node cannot be
directed to itself nor can it be part of a structure within \( B_z \) that eventually ends up being directed to itself (Figure 3.9).

\[
\begin{align*}
\text{a.} & \quad \delta_{1,z} \\
\delta_{2,z} & \quad \rightarrow \quad \delta_{3,z}
\end{align*}
\]

**Figure 3.8.** An example of a Bayesian network with three nodes. \( \delta_{1,z} \) has no parents and is the parent of \( \delta_{2,z} \) and \( \delta_{3,z} \). \( \delta_{2,z} \) has one parent, \( \delta_{1,z} \), and is the parent of \( \delta_{3,z} \). \( \delta_{3,z} \) has two parents, \( \delta_{1,z} \) and \( \delta_{2,z} \), and no children.

\[
\begin{align*}
\text{a.} & \quad \delta_{1,z} \\
\delta_{2,z} & \quad \rightarrow \quad \delta_{3,z}
\end{align*}
\]

**Figure 3.9.** Two Bayesian network structures that are disallowed in the model: a. \( \delta_{1,z} \) has a node that is directed to itself; b. \( \delta_{1,z} \) is part of a network structure that is later directed back to itself.

In order to avoid this, this study takes advantage of the nature of the problem for which this framework is developed and puts all \( \delta_{n,z} \) into ascending order according to the time in the development schedule that each \( \delta_{n,z} \) occurs. Hence, this sequence will be:

\[
\begin{align*}
\delta_{1,z} = D_{i-1}, \quad \delta_{2,z} = C_{i-1}, \quad \delta_{3,z} = T_{i-1}, \quad \delta_{4,z} = D_{i-2}, \quad \delta_{5,z} = C_{i-2}, \quad \delta_{6,z} = T_{i-2}, \ldots \\
\ldots, \quad \delta_{3,i-2,z} = D_{i-1}, \quad \delta_{3,i-1,z} = C_{i-1}
\end{align*}
\]

where \( D_i \) is the decision variable representing the choice of the \( i^{th} \) drug, \( C_i \) is the decision variable for the choice of corporate relations strategy for the \( i^{th} \) drug, and \( T_i \) is the decision choice for the length of time to wait before development of drug \( i-1 \) is commenced. The stipulation is imposed that no edge can be directed from any \( \delta_i \) to any other \( \delta_j \) which occurs before it in this specified time order.
The procedure for constructing $B_z$ from $K_z$ is described in Figure 3.10. A score based approach is used where the Bayesian Dirichlet (BD) metric and the network is constructed with the intent to maximise this score. The BD metric is a measure of how closely $B_z$ models the data in $K_z$. The construction procedure begins with an empty network and each possibility for adding the next directed edge in $B_z$ is scored. The next directed edge to be added will be the one that maximises the value of the BD metric. This process continues until no improvement in score can be achieved. Additionally, no restrictions are imposed on the complexity of the network that can be built so that no important topological details are lost.

**Bayesian network construction**

Each decision variable $\delta$ represents a node in the network to be built for the $z^{th}$ cluster, $B_z$.

Set $B_z$ as having no directed edges.

Repeat until no improvement in $p(K_z \mid B_z, \xi)$ can be made:

For all possibilities of adding the next directed edge:

Add the next directed edge.

Calculate $p(K_z \mid B_z, \xi)$.

Remove the previously added directed edge from the network.

Add the directed edge that results in the greatest value of $p(K_z \mid B_z, \xi)$.

**Figure 3.10.** Pseudocode for Bayesian network construction.

The BD metric for $B_z$ given $K_z$ and background information $\xi$, is denoted by $p(K_z, B_z \mid \xi)$ and is defined as:

$$p(K_z, B_z \mid \xi) = p(B_z \mid \xi) \prod_{n=0}^{n-1} \left( \prod_{\delta_{n,i}} \frac{\Gamma(e'(\pi_{\delta_{n,i}}))}{\Gamma(e'(\pi_{\delta_{n,i}}) + e(\pi_{\delta_{n,i}}))} \right)$$

where $p(B_z | \xi)$ is the prior probability of $B_z$; $\delta_{n,i,m}$ is the $n^{th}$ node instantiated to the $m^{th}$ value in $K_z$; the product over $\pi_{\delta_{n,i}}$ runs over all instances of all parents of $\delta_{n,i}$; the product over $\delta_{n,i,m}$ runs over all instances of $\delta_{n,i}$; $e(\pi_{\delta_{n,i}})$ is the number of instances in $K_z$ where the parent nodes of $\delta_{n,i}$, $\Pi_{\delta_{n,i}}$, are instantiated to $\pi_{\delta_{n,i}}$; $e(\delta_{n,i,m}, \pi_{\delta_{n,i}})$ is
the number of instances in $K_z$ that have $\delta_{n,z}$ equal to $\delta_{n,z,m}$ and $\Pi_{\delta_{n,z}}$ set to $\pi_{\delta_{n,z}}$; and $\Gamma(x)$ is the Gamma function where $\Gamma(x) = (x-1)!$. When the set $\Pi_{\delta_{n,z}}$ is empty there is one instance of $\Pi_{\delta_{n,z}}$ that is equal to 0, and the number of instances is set to the number of members of $K_z$, $|K_z|$. Additionally:

$$\epsilon(\pi_{\delta_{n,z}}) = \sum_{\delta_{n,z,m}} \epsilon(\delta_{n,z,m}, \pi_{\delta_{n,z}})$$  \hspace{1cm} (3.73)$$

where $\pi_{\delta_{n,z,m}}$ is an instance of the parents of $\delta_{n,z}$ that is summed over all instances of $\delta_{n,z}$. The numbers $\epsilon'(\delta_{n,z,m}, \pi_{\delta_{n,z}})$ and $p(B_z | \xi)$ represent prior information about the problem that can be incorporated into the metric.

The extent to which the network being measured represents another network relevant to the problem is measured by $p(B_z | \xi)$. There are a number of methods available for calculating $p(B_z | \xi)$. In this study all networks are treated equally thus $p(B_z | \xi)$ is set to 1. There are also a variety of methods for setting the numbers $\epsilon'(\delta_{n,z,m}, \pi_{\delta_{n,z}})$ (Buntine, 1991; Heckerman, Geiger, and Chickering, 1995; Yang and Chang, 2002).

In particular, research by Yang and Chang (2002) has demonstrated that for a range of integer values of $\epsilon'(\delta_{n,z,m}, \pi_{\delta_{n,z}})$ between 1 and 10, and also for other methods considered for scoring Bayesian network topology, the BD metric with $\epsilon'(\delta_{n,z,m}, \pi_{\delta_{n,z}}) = 10$ ranked as one of the best for discovering the true structure of the Bayesian network. Accordingly, $\epsilon'(\delta_{n,z,m}, \pi_{\delta_{n,z}}) = 10$ is used here. Additionally, because the factorials in Equation (3.72) can grow to unmanageably large numbers, especially within large sizes of $K_z$, its logarithmic equivalent is used.

3.2.2.5. Regenerating the population

In generating $G(t+1)$ each $\delta_{n,z}$ is treated as a random variable that can be instantiated to any of its possible instances subject to probabilities encoded by $B_z$. For each decision variable all probabilities for each possible $\delta_{n,z,m}$ conditional upon each
possible \( \pi_{\delta_z} \) given \( K_z \) are computationally determined and referenced when needed.

The process for generating one candidate solution for \( G(t+1) \) is shown in Figure 3.11.

<table>
<thead>
<tr>
<th>Generation of one candidate solution in ( G(t+1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>For all ( n ):</td>
</tr>
<tr>
<td>Mark all ( \delta_{n,z} ) as unprocessed.</td>
</tr>
<tr>
<td>Select an unprocessed ( \delta_{n,z} ) with all parents processed already, if any.</td>
</tr>
<tr>
<td>Randomly instantiate ( \delta_{n,z} ) to ( \delta_{n,z,m} ) with probability ( p(\delta_{n,z} = \delta_{n,z,m}</td>
</tr>
<tr>
<td>Mark ( \delta_{n,z} ) as processed.</td>
</tr>
</tbody>
</table>

**Figure 3.11.** Pseudocode for generating one candidate solution, g, for the new population \( G(t+1) \) from \( B_z \).

The joint distribution encoded by \( B_z \) can be written as:

\[
p(\delta_z) = \prod_{n=1}^{N} p(\delta_{n,z} | \Pi_{\delta_z})
\]

where \( \delta_z = (\delta_{1,z}, \delta_{2,z}, \ldots, \delta_{N,z}) \) is a vector of random variables representing the entire set of decision variables; and \( p(\delta_{n,z} | \Pi_{\delta_z}) \) is the conditional probability of \( \delta_{n,z} \) given its set of parent variables \( \Pi_{\delta_z} \). Intuitively, this conditional probability is defined as:

\[
p(\delta_{n,z} | \Pi_{\delta_z}) = \frac{p(\delta_{n,z}, \Pi_{\delta_z})}{p(\Pi_{\delta_z})}
\]

(3.75)

Each \( K_z \) is proportionally represented in the generation of \( G(t+1) \). As \( S(t) \) contains half the solutions present in \( G(t) \), the number of solutions generated using each \( B_z \) is \( 2^{|K_z|} \).

103
3.3. CASE STUDY DESCRIPTION

A hypothetical case was formulated to illustrate and examine ability of the framework to discover optimal strategies for performance against multiple objectives in an uncertain environment. In this case a biopharmaceutical company has 10 monoclonal antibody drug candidates available for development but can only choose 5. It needs to know which drug candidates should be chosen, their order of their development, the timing schedule of development activities, and which corporate bodies should be assigned to each development activity. The optimisation model is considerate of the commercial characteristics of drug candidates (Table 3.2), technical probabilities of success for each drug group (Table 3.3), durations and costs associated with various stages of the drug development (Table 3.4). As seen in Table 3.2 there are three groups of indication that these candidates belong to. Annual demand figures and compound annual growth rates (CAGR) are based on levels on realistic figures. The dependencies for revenue and capital expense are detailed in Table 3.5. Specifications for contractual dependencies and technical probabilities of success are displayed in Table 3.6.

An example of the complexity of the decision space concerning the 5 drug portfolio is discussed. Overall, each strategy consists of 54 decision variables: 5 for the selection of drugs for the portfolio, 4 for timing the commencement of development activities for the next drug in the pipeline, and 9 assignments of corporate bodies to critical activities for each of the five drugs. Hence, the choice of portfolio structure has $10 \times 9 \times 8 \times 7 \times 6$ possibilities that are combined with $^5P_3$ possibilities for the order of development. There are $7^4$ possibilities for the overall timing strategy. Additionally, there are $207^5$ possibilities for assigning corporate bodies to critical activities during the developmental phases and commercial life of each drug. Overall the entire decision space has a total of $\sim 3.31 \times 10^{31}$ individual strategies. If each strategy took 1 second to evaluate it would take $1.05 \times 10^{14}$ years to enumerate all possibilities. Hence, the combinatorial optimisation algorithm has been specifically developed to efficiently search this vast decision space.
STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT

The settings used for the mechanics of the optimisation procedure are now stated and are true for all generations. The maximum number of generations $t_{\text{max}} = 17$. The size of a given generation, $|G(t)| = 1000$. The number of superior strategies selected from a given generation, $|S(t)| = 500$. The number of clusters used in the objective space, $z = 3$.

A few comments to consider. The number of generations over which the populations are improved was tested through multiple evaluations and was indicated by the lack of progression in subsequent generations and the consistency of results across multiple tests. Because of the hierarchical super-structure of a candidate solution the total number of decisions is 14 instead of 54, which would have been the case if decisions on internalising or externalising critical activities were not grouped into a single decision. Using 250 Monte Carlo trials per candidate strategy was found to be adequate for purposes of making comparisons of quality between candidate strategies. To assure that performance metrics are entirely dependent on the strategy as opposed to being assisted by the generation of an opportunistic set of random numbers, the same set of random numbers is used across strategies. The pseudo-random number generator used here is the Mersenne Twister algorithm (Matsumoto and Nishimura, 1998). To enhance the efficiency in estimating stochastic output properties stratified sampling was used for each stochastic variable where boundaries of significance were set individually. Also, a discount rate of 20% was used for calculating the NPV values.

Table 3.2. Commercial characteristics of available drug candidates.

<table>
<thead>
<tr>
<th>Drug Candidate</th>
<th>Drug Group</th>
<th>Annual Demand (kg/year)</th>
<th>CAGR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1</td>
<td>Tr$(120,250,380)$</td>
<td>Tr$(1.00,1.01,1.02)$</td>
</tr>
<tr>
<td>$B$</td>
<td>1</td>
<td>Tr$(100,200,300)$</td>
<td>Tr$(1.00,1.01,1.02)$</td>
</tr>
<tr>
<td>$C$</td>
<td>1</td>
<td>Tr$(80,150,230)$</td>
<td>Tr$(1.00,1.01,1.02)$</td>
</tr>
<tr>
<td>$D$</td>
<td>2</td>
<td>Tr$(50,100,150)$</td>
<td>Tr$(1.00,1.01,1.02)$</td>
</tr>
<tr>
<td>$E$</td>
<td>2</td>
<td>Tr$(100,200,300)$</td>
<td>Tr$(1.01,1.03,1.04)$</td>
</tr>
<tr>
<td>$F$</td>
<td>2</td>
<td>Tr$(80,150,230)$</td>
<td>Tr$(1.01,1.03,1.04)$</td>
</tr>
<tr>
<td>$G$</td>
<td>2</td>
<td>Tr$(50,100,150)$</td>
<td>Tr$(1.01,1.03,1.04)$</td>
</tr>
<tr>
<td>$H$</td>
<td>3</td>
<td>Tr$(80,150,230)$</td>
<td>Tr$(1.02,1.05,1.06)$</td>
</tr>
<tr>
<td>$I$</td>
<td>3</td>
<td>Tr$(50,100,150)$</td>
<td>Tr$(1.02,1.05,1.06)$</td>
</tr>
<tr>
<td>$J$</td>
<td>3</td>
<td>Tr$(100,200,300)$</td>
<td>Tr$(1.02,1.05,1.06)$</td>
</tr>
</tbody>
</table>
Table 3.3. Duration and cost information for various phases of the drug development process.

<table>
<thead>
<tr>
<th>Phase of Development</th>
<th>Duration (years)</th>
<th>Cost ($MM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Identification</td>
<td>1</td>
<td>Tr(3,5,8)</td>
</tr>
<tr>
<td>PreClinical</td>
<td>2</td>
<td>Tr(20,35,50)</td>
</tr>
<tr>
<td>Phase I clinical trials</td>
<td>1</td>
<td>Tr(5,15,20)</td>
</tr>
<tr>
<td>Phase II clinical trials</td>
<td>2</td>
<td>Tr(15,25,35)</td>
</tr>
<tr>
<td>Phase III clinical trials</td>
<td>3</td>
<td>Tr(45,85,125)</td>
</tr>
<tr>
<td>Scale-up synthesis</td>
<td>1</td>
<td>Tr(3,5,8)</td>
</tr>
<tr>
<td>Formulation</td>
<td>1</td>
<td>Tr(5,10,15)</td>
</tr>
<tr>
<td>Commercial Preparation</td>
<td>1</td>
<td>Tr(1,2,3)</td>
</tr>
<tr>
<td>Marketing</td>
<td>1</td>
<td>Tr(2,4,6)</td>
</tr>
<tr>
<td>FDA Review</td>
<td>1</td>
<td>Tr(2,4,6)</td>
</tr>
<tr>
<td>Market Lifetime</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>Ramp time to peak market: 'I'</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
<tr>
<td>Ramp time to peak market: 'C'</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
<tr>
<td>Ramp time to peak market: 'P'</td>
<td>Tr(2,1,3)</td>
<td>-</td>
</tr>
<tr>
<td>Decay time after market expiry: 'I'</td>
<td>Tr(2,1,3)</td>
<td>-</td>
</tr>
<tr>
<td>Decay time after market expiry: 'C'</td>
<td>Tr(2,1,3)</td>
<td>-</td>
</tr>
<tr>
<td>Decay time after market expiry: 'P'</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: The cost of producing a drug for the marketing phase is determined by the biomanufacturing cost model. '-' means that particular combination of referencing items is irrelevant.

Table 3.4. Technical information for manufacture of drug candidates according to clinical phase and corporate body.

<table>
<thead>
<tr>
<th>Corporate Body</th>
<th>Developmental Phase</th>
<th>Whole Process Yield (%)</th>
<th>Fermentation Titre (g/L)</th>
<th>Batch Success Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>'I'</td>
<td>Phase I</td>
<td>Tr(0.25,0.35,0.50)</td>
<td>Tr(0.20,0.30,0.50)</td>
<td>Tr(0.8,0.9,1.00)</td>
</tr>
<tr>
<td>'C'</td>
<td>Phase I</td>
<td>Tr(0.45,0.55,0.85)</td>
<td>Tr(0.8,1.00,1.50)</td>
<td>Tr(0.7,0.9,1.00)</td>
</tr>
<tr>
<td>'P'</td>
<td>Phase I</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.60,0.75,1.20)</td>
<td>Tr(0.65,0.85,1.00)</td>
</tr>
<tr>
<td>'I'</td>
<td>Phase II</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.30,0.40,0.60)</td>
<td>Tr(0.65,0.85,1.00)</td>
</tr>
<tr>
<td>'C'</td>
<td>Phase II</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
<tr>
<td>'P'</td>
<td>Phase II</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>'I'</td>
<td>Phase III</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>'C'</td>
<td>Phase III</td>
<td>Tr(0.55,0.75,1.00)</td>
<td>Tr(1.50,2.00,3.00)</td>
<td>Tr(0.75,1.00,1.00)</td>
</tr>
<tr>
<td>'P'</td>
<td>Phase III</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
<tr>
<td>'I'</td>
<td>Market</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>'C'</td>
<td>Market</td>
<td>Tr(0.55,0.70,1.00)</td>
<td>Tr(1.50,2.00,3.00)</td>
<td>Tr(0.75,1.00,1.00)</td>
</tr>
<tr>
<td>'P'</td>
<td>Market</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
</tbody>
</table>
Table 3.5. Specification of dependencies as related to the number of drugs from the same group within the chosen drug development portfolio.

<table>
<thead>
<tr>
<th>Number of drugs</th>
<th>% of full revenue</th>
<th>% of full capital</th>
<th>% of full COGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>85%</td>
<td>93%</td>
<td>95%</td>
</tr>
<tr>
<td>3</td>
<td>75%</td>
<td>85%</td>
<td>90%</td>
</tr>
<tr>
<td>4</td>
<td>65%</td>
<td>78%</td>
<td>85%</td>
</tr>
</tbody>
</table>

Note: COGS = cost of goods sold.

Table 3.6. Specification of contractual dependencies and stage-wise probabilities of success.

<table>
<thead>
<tr>
<th>Development Stage</th>
<th>Royalty rate (% of revenue)</th>
<th>Probability of success</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CRO</td>
<td>CMO</td>
</tr>
<tr>
<td>Target identification</td>
<td>6%</td>
<td>-</td>
</tr>
<tr>
<td>Pre-clinical trials</td>
<td>7%</td>
<td>-</td>
</tr>
<tr>
<td>Phase I clinical trials</td>
<td>8%</td>
<td>10%</td>
</tr>
<tr>
<td>Phase II clinical trials</td>
<td>9%</td>
<td>12%</td>
</tr>
<tr>
<td>Phase III clinical trials</td>
<td>10%</td>
<td>14%</td>
</tr>
<tr>
<td>Market</td>
<td>-</td>
<td>16%</td>
</tr>
</tbody>
</table>

Note: A decrease in royalty rate applies per additional drug and is measured in percentage points. Royalty rates decrease by 1% with the CRO, and 2% with the CMO or partner for each additional drug development project for which they are involved. A minimum royalty rate applies of 3% with the CRO, 5% with the CMO, and 40% with the partner. Groups 1, 2, and 3 refer to drug groups 1, 2, and 3. '-' means that particular combination of referencing items is irrelevant.

3.4. RESULTS

The case study results are discussed in the following sections focusing on analysing competing strategies along the Pareto front generated and identifying trends along the frontier using cluster analysis.

3.4.1. PARETO FRONT PROGRESSION

In this section the performance of strategies in the final population is examined in terms of their ability to generate profits, satisfy multiple objectives, and provide an acceptable risk profile. The analysis highlights the challenges that exist when pursuing multiple objectives in the context considered.
The progression of the EDA in discovering the Pareto front is demonstrated in Figure 3.12. As might be expected, random initialisation of the population of strategies has resulted in a generally even dispersion of the mean positive NPV and $p(\text{NPV}>0)$ performance attributes over the objective space. Interestingly, what can immediately be seen is that it is possible to construct a portfolio of drugs with a significantly higher probability of attaining a profit than any of its individual comprising drugs. Here drugs with a probability of reaching market that is less than 0.25 have been pooled in such a way that offers the decision maker a $p(\text{NPV}>0)$ value in excess of 0.70. Progression of the algorithm in the objective space appears to be considerably more constrained in the $p(\text{NPV}>0)$ dimension than it is in the mean positive NPV dimension. Progression of the evolutionary process has been accompanied by some degree of succession at each generation. It is apparent that the strategies available in the objective space have generally converged towards a clearly defined non-
dominated frontier, the Pareto front. The Pareto front is wide enough to be inclusive of \( p(\text{NPV}>0) \) values that might be realistically sought after by a decision maker in this setting. It is also apparent that for this problem the machine learning mechanisms have greatly reduced the variability encoded in the conditional and probabilistic models of superior strategies. It should be noted that the ability of the algorithm to do this is dependent on whether a new set of building blocks is able to clearly distinguish itself as superior to any existing alternatives in the evolutionary process. At some stages, succession in some regions of the objective space is accompanied by deterioration in other regions. This deterioration refers specifically to the non-population of certain regions that were previously populated. In this particular case, it was considered appropriate that the convergence of the algorithm was tracked visually, principally because the objective space has only two dimensions. The optimisation procedure was considered to be converged when general progression of the Pareto front was insignificant, leading to an unfavourable loss of regions along the Pareto front. Some deterioration was tolerated in regions where rational decision makers were unlikely to seek strategies. An example of such a region would be where \( p(\text{NPV}>0) \) is less than 0.20 or where the mean positive NPV is less than $200M. Further analysis indicated that progression beyond 17 generations for this case study leads to a severe deterioration of the Pareto front.

The superior strategies selected from the seventeenth generation (Figure 3.13) show that a non-dominated frontier exists in the approximate region 0.18<\( p(\text{NPV}>0) \)<0.75. The frontier indicates that a definite trade-off exists here between maximizing the mean positive NPV to be generated by a particular strategy and maximizing the probability of attaining a positively valued profit, \( p(\text{NPV}>0) \). At no point along the frontier can the maximisation of mean positive NPV and \( p(\text{NPV}>0) \) be aligned. Hence, improving \( p(\text{NPV}>0) \) means accepting a degradation in mean positive NPV. Also, this forces the decision maker into the process of selecting strategies that begins with deciding upon minimum acceptable levels of profit and probability of profitability. The region 0.50<\( p(\text{NPV}>0) \)<0.65 offers the least trade-off along the frontier when searching for strategies that improve \( p(\text{NPV}>0) \). In practical terms, not all of the frontier will be appealing to the decision maker as there will be strategies
that either offer probabilities of success or mean positive NPV values that are too low to be considered for further action.

Figure 3.13. Mean positive NPV versus p(NPV>0). Note: (●) represents an estimate of the non-dominated frontier. Strategies chosen for further analysis are annotated as S1, S2, S3, and S4.

To assess the effectiveness of the EDA, the Pareto front discovered is compared with that of a simple random search (Figure 3.14). For the random search seventeen thousand unique strategies were randomly generated and then evaluated. The rationale for generating this number of strategies for the random search in this case is based on the EDA evaluating seventeen thousand strategies before termination. It can be seen that the EDA offers a marginal improvement above the performance obtained from random search. It is also observed that at the extremities of the objective space there is negligible difference in the performance offered by either algorithm. This indicates that either a performance limit has been reached or that the EDA is simply ineffective for these regions. Without an algorithm that guarantees discovery of the true Pareto front it is impossible to identify which postulation is correct. Certainly the fact that these extremities were discovered by random search suggests that they were relatively easy to find. Where the EDA performs more convincingly above random search is in the intermediary region between the extremities. Across the Pareto front the average improvement afforded by the EDA is approximately 10% in mean positive NPV for a given p(NPV>0). The improvement seen can be up to 20%.
STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT

Figure 3.14. Pareto fronts for Mean positive NPV versus p(NPV>0). Note: (—) is the Pareto front discovered by the EDA, and (---) is the Pareto front discovered by random search.

3.4.2. ANALYSIS OF COMPETING STRATEGIES

A close examination of competing strategies along the frontier indicates that they can have similar reward-risk characteristics whilst demonstrating marked differences in key decisions relating either to the portfolio structure (drug selection), timing or third party strategies (Table 3.7). Four strategies along the non-dominated frontier have been selected to illustrate this and demonstrate how a decision maker might rationalise the choice between two strategies of similar risk and reward as visualised in Figure 3.13.

Table 3.7. Attributes of strategies chosen for comparison.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>p(NPV&gt;0)</th>
<th>Positive NPV ($MM)</th>
<th>Negative NPV ($MM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Min.</td>
</tr>
<tr>
<td>SI</td>
<td>0.184</td>
<td>$1421</td>
<td>$17</td>
</tr>
<tr>
<td>S2</td>
<td>0.192</td>
<td>$1420</td>
<td>$10</td>
</tr>
<tr>
<td>S3</td>
<td>0.368</td>
<td>$1126</td>
<td>$12</td>
</tr>
<tr>
<td>S4</td>
<td>0.376</td>
<td>$1082</td>
<td>$18</td>
</tr>
</tbody>
</table>

Note: Min. Minimum, Max. Maximum, SSD semi-standard deviation.

SI and S2 are strategies that generate the greatest mean positive financial reward but also the greatest risk. Although these competing strategies have very similar third
party strategies and portfolio structures, they differ greatly in their timing strategies as illustrated in Table 3.8 and Table 3.9. Both strategies share the approach of partnering for the majority of activities for the first drug and then preferring to use a combination of in-house and outsourced development and manufacturing for the remaining drugs. The only exception to this is for drug 3 with SI where partnering is used for commercial manufacturing. The first two drugs are either developed within three years of each other, as in SI, or simultaneously, as in S2, and developing the remaining drugs once these drugs are close to the marketplace. This approach has the advantage of dividing the risk and impact of failure between two drug development projects and it is clear that the expected profits from either of the first two drugs are used to fund the development of future projects. SI staggers the development of its drugs with lengthy development intervals between each drug. Resultantly the time it takes to complete development of the portfolio is 35 years. S2 has a shorter development time of 26 years because of the shorter intervals between the developments of its final three drugs. Each strategy develops almost the same portfolio of drugs with the difference being that for the fourth drug SI develops Drug G and S2 develops Drug B. Much like the strategy with the first two drugs, SI develops its third and fourth drugs together and S2 develops its fourth and fifth drugs within three years of each other. For SI these two drugs are taken from the same group whereas with S2 this is not the case. SI and S2 generate similar performances in mean positive NPV but a strong reason for selecting S2 exists. S2 has the additional advantage of being executed over a significantly shorter period.

Table 3.8. Structure of the SI portfolio development strategy.

<table>
<thead>
<tr>
<th>i</th>
<th>D₁</th>
<th>C₁₁</th>
<th>C₁₂</th>
<th>C₁₃</th>
<th>C₁₄</th>
<th>C₁₅</th>
<th>C₁₆</th>
<th>C₁₇</th>
<th>C₁₈</th>
<th>C₁₉</th>
<th>Tᵢ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td>'I'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'FDA'</td>
</tr>
<tr>
<td>3</td>
<td>J</td>
<td>'C'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'FDA'</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>'I'</td>
<td>'I'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'II'</td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>'C'</td>
<td>'T'</td>
<td>'I'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
</tr>
</tbody>
</table>

Note: 'I' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, C₁₁ - target identification; C₁₂ - preclinical studies; C₁₃ - Phase I clinical development; C₁₄ - Phase II clinical development; C₁₅ - Phase III clinical development; C₁₆ - Phase I manufacturing; C₁₇ - Phase II manufacturing; C₁₈ - Phase III manufacturing; C₁₉ - Market, '-' - that particular combination of referencing items is irrelevant.

112
Examining the differences between another pair of competing strategies illustrates how marked differences in third party strategies can also result in similar reward-risk characteristics. In contrast to S1 and S2, strategies S3 (Table 3.10) and S4 (Table 3.11) possess similar timing strategies but very different third party strategies as well as portfolio structures. S3 utilises a mixture of in-house, outsourced, and partnered activities whereas S4 develops the entire portfolio in-house. The portfolio constructed by S3 consists of one low demand drug, two medium demand drugs, and two high demand drugs that are sourced from all three groups. The first three drugs are developed first and simultaneously with the remaining two drugs developed with lengthy intervals of nine years in between. For the second and third drugs contractors are only used for clinical development, clinical manufacturing is kept in-house, and commercial manufacturing is conducted with a partner. The fourth and fifth drugs use contracting more intensively for clinical development but contractors are also used for clinical manufacturing where they are used during phase III trials. The total portfolio development time plus time taken during the marketing phase for S3 is 28 years. S4 develops drugs from all three groups and its portfolio consists of two low demand drugs, two medium demand drugs, and one high demand drug. Similar to S3, the first three drugs are developed in close succession and also stagger the development of its final two drugs in a similar fashion. The total time for development and marketing completion of S4 is 30 years, which is similar to that taken for S3. Comparing their performances, S3 generates a marginally superior mean positive NPV than S4 and is
executed over a slightly shorter period. Although marginal, the performance results suggest S3 as being the superior strategy.

Table 3.10. Structure of the S3 portfolio development strategy.

<table>
<thead>
<tr>
<th>i</th>
<th>Di</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>Ti</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>T</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>'C'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'P'</td>
<td>'FDA'</td>
</tr>
<tr>
<td>5</td>
<td>J</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
</tr>
</tbody>
</table>

Note: 'I' – in-house activity, 'C' – outsourced activity, 'P' – partnered activity, C1 – target identification; C2 – preclinical studies; C3 – Phase I clinical development; C4 – Phase II clinical development; C5 – Phase III clinical development; C6 – Phase I manufacturing; C7 – Phase II manufacturing; C8 – Phase III manufacturing; C9 – Market, ‘-’ – that particular combination of referencing items is irrelevant.

Table 3.11. Structure of the S4 portfolio development strategy.

<table>
<thead>
<tr>
<th>i</th>
<th>Di</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>Ti</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>T</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'PC'</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'I'D'</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'FDA'</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>J</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'FDA'</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
</tr>
</tbody>
</table>

Note: 'I' – in-house activity, 'C' – outsourced activity, 'P' – partnered activity, C1 – target identification; C2 – preclinical studies; C3 – Phase I clinical development; C4 – Phase II clinical development; C5 – Phase III clinical development; C6 – Phase I manufacturing; C7 – Phase II manufacturing; C8 – Phase III manufacturing; C9 – Market, ‘-’ – that particular combination of referencing items is irrelevant.

From sampling the strategies along the Pareto front it is observable that their positioning in the objective space and even their occurrence within the optimised population of strategies is not intuitive. It is taken that this is due to the complexity that is inherent in the model of the biopharmaceutical drug development pathway and the complexity that governs the formulation of superior strategies. This is an important underscore because it highlights to the decision maker the value of accounting for the concept of building blocks, which is considerably difficult to achieve without the use of advanced computational tools. Another important
observation is that strategies with clear differences in either drug selection, timing, or third party strategies can compete with similar reward versus risk profiles. Hence it is useful for the decision-maker to identify a desirable region along the frontier and closely examine the different options that can yield the desired return and acceptable risk.

### 3.4.3. Cluster Analysis

In order to investigate if any discernable trends in strategy formulation exist amongst the population of superior strategies, the objective space was decomposed into three clusters. It was anticipated that an analysis of these clusters might add useful insight when considering if any particular regions of the Pareto front give rise to the prevalence of certain building blocks. As building blocks effectively compete with each other for selection it would be useful to discover if and why any particular building blocks are emphasised.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Characteristic</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
<th>Drug 4</th>
<th>Drug 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demand</td>
<td>M</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>Drug Group</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Timing</td>
<td>'ID'</td>
<td>'FDA'</td>
<td>'PIII'</td>
<td>'FDA'</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Demand</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>Drug Group</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Timing</td>
<td>'PC'</td>
<td>'PC'</td>
<td>'MKT'</td>
<td>'FDA'</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Demand</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Drug Group</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Timing</td>
<td>'ID'</td>
<td>'FDA'</td>
<td>'PII'</td>
<td>'PII'</td>
<td>-</td>
</tr>
</tbody>
</table>

**Portfolio structure**

Table 3.12 displays the most probable constituents of strategy for each cluster. This includes the annual level of demand for the selected drug, the group that the drug belonged to, and the time after which development of the next drug would commence. The reader is reminded that group 1 drugs have a low annual growth rate, while groups 2 and 3 respectively have medium and high annual growth rates. Also, group
3 drugs have the greatest probability of achieving marketing approval whilst group 2 drugs have the least. Cluster 1 consists of strategies that perform highly in mean positive NPV, such as $S_1$ and $S_2$, and cluster 3 consists of strategies that have high values of $p(\text{NPV}>0)$. Cluster 2 represents strategies that balance both of these performance metrics more evenly, such as $S_3$ and $S_4$. The clusters and their member strategies are derived by application of the k-means clustering algorithm. For cluster 1, the most probable portfolio consists of one low demand drug, one medium demand drug and three high demand drugs from all three drug groups. Strategies in this cluster tend to utilise the largest number of high demand drugs which assists in explaining the larger mean positive NPV values generated by these strategies. Also, a larger portion of selected drugs belong to drugs with high and medium probabilities of success. For cluster 3 at the opposite end of the Pareto front, the most probable portfolio is three medium demand drugs and two high demand drugs drawn from all three drug groups, which perhaps surprisingly highlights that cluster 3 strategies aim to meet a similar level of market demand as cluster 1 drugs. As cluster 3 strategies have significantly lower mean positive NPV values than cluster 1 strategies this indicates that there are other decisions outside of portfolio structure which lower this profitability measure but maybe contribute to a greater $p(\text{NPV}>0)$ value. With the exception of one drug the portfolio opts for drugs with high and medium probabilities of success thereby using more drugs of this type than the other clusters. This is intuitive because this cluster exhibits the greatest $p(\text{NPV}>0)$ values. Overall, there appear to be no clear trends in portfolio structuring strategies. The presence of certain portfolio structuring decisions within certain clusters appears to align with the position in the objective space relative to other clusters, while there are others which do not. The remaining constituents of the portfolio development strategy will be investigated to discover if more overarching drivers of performance exist.

**Timing strategies**

Decisions on timing are an important constituent of the portfolio development strategy as they are used to favourably organise cash flows. This is particularly important when having to consider the probability that a project will succeed, the financial impact of failed projects, and the impact of the discount factor when determining the mean positive NPV. In cluster 1, Table 3.12 shows that the prevailing strategy is to develop the first two drugs together coupled with long
intervals of either eight or nine years between the development of the remaining drugs. Because of the probabilities of success of available drugs it would be rational to expect only one successful drug within a portfolio of five projects. Also, it is observable that when considering the growth and impact of the discount factor over time, cluster 1 strategies must rely considerably on the production of a successful drug from the first two projects. This assists in reasoning the low $p(\text{NPV}>0)$ values seen in this cluster. Remaining drugs have comparatively little impact on cash flows because of the extent to which their respective cash flows are discounted. This also provides an explanation for the superior mean positive NPV values as when one successful drug emerges from within the first two projects its profits are not greatly discounted. Also and importantly, profits from this drug will mainly need to absorb the expense of its own development and that of the other concurrent project. The cost of developing the remaining drugs is less restrictive as these are heavily discounted. Albeit this occurs with a relatively modest probability. The time for development until the end of marketing for this strategy is 34 years. For cluster 2 it is shown that the first three drugs are most likely to be developed with short intervals of one year between them and that the time for development and marketing completion for this approach is 31 years. This strategy is a driver of the higher $p(\text{NPV}>0)$ values and the lower mean positive NPV values observed when making comparison to cluster 1. Like cluster 1, remaining drugs are likely to be developed with the longest intervals between them and are anticipated to have modest impacts on cash flow because of the magnitude of the discount factor at these stages in portfolio development. Considering this, such strategies are reliant on at least one successful drug emerging from the first three projects. Developing the first three drugs in close succession in cluster 2 bears a greater likelihood that at least one successful project will emerge from this group than from within the remaining two drugs. When a successful drug emerges from this group, because of a more favourable discount factor there is also a greater likelihood that it can also cover the expense of other failed projects than when developing two drugs in close succession. Having to effectively absorb the cost of an increased number of projects whose development costs are less discounted contributes to lower mean positive NPV values. Strategies in cluster 3 exhibit a tendency to develop the first two drugs together followed by a period of at least nine years then later drugs are developed with medium length intervals of four years in between them. The time for development and marketing completion most likely in cluster 3 is 27 years. This
strategy relies on at least one successful drug being found in two distinct groups, that is a group of two drugs developed together at the outset and a group of three drugs developed in close succession later in the timeline. It has already been mentioned that the impact of developing two drugs together whilst positioning the remaining projects to subject them to far more significant discounting results in relatively high mean positive NPV values with relatively low probabilities of success. It follows that some aspects of the development of the remaining three drugs contribute significantly to the high \( p(\text{NPV}>0) \) values but low mean positive NPV values ultimately exhibited by strategies in this cluster. The shorter portfolio development time means that profits for each year are discounted by a significantly smaller extent than seen with clusters 1 and 2. Strategies in this cluster also take advantage of the high annual market growth rate and high probability of success of group 3 drugs and position them late in the pipeline of products to maximise their potential for generating revenues that overcome the magnitude of the discounting. This is important as if each drug is to contribute significantly to the revenue generated by the portfolio then any successful drugs must support up to four failed projects, which helps to explain the smaller mean positive NPV values. Also it is more likely to see at least one successful drug emerge from within the second group of three projects than from the first group of two. When this occurs, profits arising from these drugs which are heavily discounted and must also be enough to cover the expenses of failed projects which are less heavily discounted. This further erodes the potential magnitude for mean positive NPV. Overall, the trend with timing strategies is that portfolio development times tend to become progressively shorter towards the right hand side of the frontier and that strategies make use of grouping projects more closely together. Interestingly, the presence of the discount factor appears to have a significant influence on strategy formulation. Hence it is anticipated that alternative settings for the discount rate may result in a noticeably different set of superior strategies.
### Table 3.13. Selected probabilities of third party strategies.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Drug</th>
<th>$D_1$</th>
<th>$C_{i1}$</th>
<th>$C_{i2}$</th>
<th>$C_{i3}$</th>
<th>$C_{i4}$</th>
<th>$C_{i5}$</th>
<th>$C_{i6}$</th>
<th>$C_{i7}$</th>
<th>$C_{i8}$</th>
<th>$C_{i9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>Drug 1</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>P'</td>
<td>T'</td>
<td>T'</td>
<td>P'</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 2</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>P'</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 3</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>'C'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>P'</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Drug 4</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>T'</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 5</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>T'</td>
<td>C'</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Drug 1</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>P'</td>
<td>T'</td>
<td>P'</td>
<td>P'</td>
<td>P'</td>
<td>P'</td>
<td>P'</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 2</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>P'</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 3</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>T</td>
<td>T</td>
<td>P'</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 4</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 5</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>P'</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Drug 1</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>P'</td>
<td>T'</td>
<td>T'</td>
<td>P'</td>
<td>P'</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 2</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>T</td>
<td>T</td>
<td>C'</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 3</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>'C'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>T</td>
<td>T</td>
<td>C'</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Drug 4</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>T</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug 5</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

*Note: For each $D_i$ in each cluster the two most probable strategies have been displayed.*

The probability of selecting a particular strategy for a particular drug position in a particular cluster of the objective space is addressed and the top two strategies in each
circumstance are shown in Table 3.13. In all cases the most probable third party strategy for each drug is to integrate all activities in-house. This is intuitive from the viewpoint of minimizing contracting fees and premiums as well as avoiding the sharing of sales revenues and royalty charges. As seen in Table 3.13 and with strategies $S1$ through $S4$, alternative third party strategies are present in significant proportion within the population. This demonstrates that when supported by appropriate strategic decisions in other areas these are viable constituents of superior strategies. It is postulated that one possible reason for complete in-house strategies to be present with significant likelihoods is that such strategies may serve as a more flexible building block than others for creating superior portfolio development strategies. With the aim of managing the impact of failure it may seem counter-intuitive for a company to go the entire drug development process without third party assistance. Because no budgetary constraint exists, in such scenarios there is no direct requirement to limit the impact of failure. However, to some degree maximizing mean positive NPV is likely to involve some mitigation of the impact of failure. As might be expected, it can be seen that strategies in cluster 1 are most likely to choose to develop drugs entirely in-house as these offer the greatest potential for achieving the greatest mean positive NPV values if the drug passes all stages of testing and review. For the first drug an alternative strategy, albeit much less likely in this case, is to partner from the commencement of phase III clinical trials or from pre-clinical testing. Partnering from phase III clinical trials allows for some savings on the proportion of revenue that must be paid to the partner. Partnering from preclinical testing onwards allows for the impact of failure to be shared but a substantial share of any sales revenues must be paid to the partner. Across all clusters strategies tend to utilise partners much less intensively from drug 1 onwards, where partners are used almost exclusively for commercial manufacturing and contractors are preferred. The use of contractors does not appear to be limited to a particular set or series of activities and in most cases the use of contractors is used as a supplement to integrated activities. Finally, although the model shows that in-house activities are the most probable for optimised strategies, the presence of budgetary constraints is expected to impact such a decision. This is explored further in chapter 4.
3.5. CONCLUSIONS

The development of a stochastic multi-objective combinatorial optimisation framework has been presented that addresses three key decisions simultaneously: portfolio management, scheduling of drug development and manufacturing, and the involvement of third parties for specific activities has been presented. Demonstrated within this work is the value carried by considering these critical strategic considerations within a unified framework that simulates and optimises all such decisions across the entire product portfolio. A case study was used to illustrate the capabilities of the framework and also highlighted that the scope of decisions that a drug developer may be confronted with can be vast and complex. Due to the complexity of this problem, a principle contribution of this work is in demonstrating a formulation based on techniques from artificial intelligence, in particular evolutionary computation and machine learning, employed for an efficient search of the decision space and for effective traversing of the objective space.

It is proposed that biopharmaceutical product development strategies in the real world may be better analysed when considering the impact of decisions holistically rather than only individually. One reason for this is the presence of dependencies between decisions that may impact economic relationships. Another is that it has been demonstrated that an effective strategy for portfolio development can result in a $p(\text{NPV}>0)$ value that is significantly greater than the probability of successful development for any singular drug in the portfolio. Hence, by considering a portfolio of multiple drugs it is possible to control its risk to some extent through careful strategic formulation, whilst this is not possible with a singular drug. Use of the model has highlighted that pursuit of mean positive NPV can conflict with pursuit of high $p(\text{NPV}>0)$ values, although this may change under different case study settings. The results of the case study lean towards suggesting the integration of all activities in-house. This can conflict with common perspectives in industry that accept such strategies with reluctance because of the uncertainty of the drug development process and the consequential impact of failure. The added presence of budgetary constraints and a range of sizes for the portfolio would serve as factors that can further capture limitations in the real world and are also capable of significantly influencing results.
STOCHASTIC COMBINATORIAL OPTIMISATION APPROACH TO BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT

seen here; this is explored in chapter 4. Furthermore the optimal set of solutions is expected to be sensitive to the relative difference in manufacturing efficiencies assumed between in-house and external manufacturing.

Finally, the learning of Bayesian networks from superior solutions presented here has been shown to be effective and efficient in improving the population, and in discovering a dense and widespread Pareto front when compared to random search in the regions that are not at the extreme of either dimension of the objective space. Their effectiveness in these regions is presumed to be due to their ability to iteratively learn and exploit the structure of the problem as noted by contributions in artificial intelligence literature. At the extremity of each objective it is observed that any comparative improvement to random search here is negligible. It is not known whether this is due to a limit inherent to the problem itself or to the algorithm itself. This should be investigated in future work. A noteworthy insight from using the framework is that use of machine learning has potential for future development in solving portfolio development and capacity planning problems simultaneously.

3.6 NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_z$</td>
<td>Bayesian network pertaining to cluster $K_z$</td>
</tr>
<tr>
<td>$C_g$</td>
<td>third party strategy of the $g^{th}$ strategy in $G(t)$ whose instances are combinations of: 'I' (in-house activity), 'C' (outsourced activity), and 'P' (partnered activity).</td>
</tr>
<tr>
<td>$D_g$</td>
<td>drug development sequence of the $g^{th}$ strategy in $G(t)$</td>
</tr>
<tr>
<td>$G(t)$</td>
<td>a generation of candidate strategies at iteration $t$</td>
</tr>
<tr>
<td>$\lvert G(t) \rvert$</td>
<td>number of candidates in $G(t)$</td>
</tr>
<tr>
<td>$g$</td>
<td>$g^{th}$ candidate solution in $G(t)$, also a subscript.</td>
</tr>
<tr>
<td>$K_z$</td>
<td>$z^{th}$ cluster of candidate strategies in $S(t)$</td>
</tr>
<tr>
<td>$\lvert K_z \rvert$</td>
<td>size of $K_z$</td>
</tr>
<tr>
<td>NPV</td>
<td>net present value</td>
</tr>
<tr>
<td>$O_z(t)$</td>
<td>set of strategies generated from $B_z$</td>
</tr>
<tr>
<td>$S(t)$</td>
<td>set of superior candidate strategies in $G(t)$ at iteration $t$</td>
</tr>
<tr>
<td>$\lvert S(t) \rvert$</td>
<td>number of candidates in $S(t)$</td>
</tr>
<tr>
<td>$t_{MAX}$</td>
<td>maximum number of iterations in the optimisation procedure</td>
</tr>
</tbody>
</table>
Tg scheduling strategy of the gth strategy in G(t) whose instances are combinations of: ‘ID’ (target identification phase), ‘PC’ (pre-clinical development phase), ‘PI’ (phase I clinical trials), ‘PII’ (phase II clinical trials), ‘PIII’ (phase III clinical trials), ‘FDA’ (FDA review phase), and ‘MKT’ (marketing phase).

Greek Symbols:
- δz,m nth node in Bz of the zth cluster instantiated to the mth value
- Φ the set of Monte Carlo trials
- φ result of a single Monte Carlo trial
- Πδz,x parent set of δz,x
- πδz,x nth instance of Πδz,x
- ε(δz,m,πδz,x) number of instances in Kz where δz,m is set to δz,m and Πδz,x set to πδz,x

Subscripts:
- i ∈ I drugs in the drug development sequence
- j ∈ J development activities
- u ∈ U Monte Carlo trials
- v ∈ V objectives
- z ∈ Z clusters
CHAPTER 4

STRATEGIC BIOPHARMACEUTICAL PORTFOLIO DEVELOPMENT: AN ANALYSIS OF CONSTRAINT INDUCED IMPLICATIONS ON STRATEGIES

4.1. INTRODUCTION

The framework introduced in chapter 3 is now used for further exploration of the decision making landscape. Some of the challenges of formulating appropriate biopharmaceutical research and development strategies include deciding how many drugs should comprise the portfolio given a group of candidates, and finding a suitable strategic approach that satisfies any constraints on cash flow whilst maintaining an optimal balance between risk and reward. Accordingly, these issues are addressed here. The objective of finding optimal drug development portfolios has been covered in literature (for example see Blau et al., 2004; Rogers et al., 2002) where the most common approach is to optimise the content of the portfolio given a particular problem formulation whilst bound by some form of monetary constraint. Also, the assumption that is often taken in illustrative cases is that the company under analysis wishes to develop all drugs by themselves and usually all at once. To extend this concept, this chapter relaxes these assumptions by investigating the discovery of optimal strategies that include possibilities for assigning critical development activities to third parties in as well as an optimal yearly schedule by which these activities should occur. Optimal strategies will maximise a level of reward for a particular measure of risk which are namely the mean positive NPV and the probability of obtaining a profit, p(NPV>0), respectively. When optimised, strategies in the objective space comprised of mean positive NPV returns and p(NPV>0) will form a non-dominated frontier that is expected to present a trade-off to the decision maker. Explicitly this trade-off is likely to mean that opting to achieve greater expected profits will also entail accepting a lesser p(NPV>0). It is the aim of this paper to investigate how the construction of performance optimised strategies and the
position of this frontier in the objective space will vary according to selected constraints. These constraints are the size of the portfolio and the magnitude of the constraint on cash flow.

4.2. Method

The optimisation procedure, settings, and case study assumptions are the same as those described in chapter 3, with any additions mentioned here. The technique measures two objectives: mean positive NPV and p(NPV>0).

Additionally, the effect of constraints on portfolio size and cash flow were investigated to highlight insights from the illustrative cases, these are the only two constraints considered. The first portfolio consisted of 5 drugs and the cash flow constraints applied to this portfolio were unconstrained, -$200MM, -$100MM, and -$75MM. The second portfolio consisted of 3 drugs and the cash flow constraints applied are unconstrained, -$100MM, and -$75MM. Specifically, each cash flow constraint, $z_{i,\text{const}}$, represented the maximum negative net present value (NPV) that the company's cash flow was allowed to reach. For both sizes of portfolio, each could be constructed from the same set of available drug candidates, which are also identical to those seen in chapter 3. The optimisation model is considerate of the commercial characteristics of drug candidates, technical probabilities of success for each drug group, durations and costs associated with various stages of the drug development. As previously seen there are three groups of indication that these candidates belong to. Annual demand figures and compound annual growth rates (CAGR) are based on levels on realistic figures. The dependencies for revenue, capital expense and royalty dependencies are identical to those displayed in chapter 3. For simplicity, no outside competition from other drug developers is assumed. The stochastic variables included in this work are exclusively characterised by way of triangular probability distributions because of their convenience when limited sample data are available. For instance, if a Gaussian distribution was used a mean and standard deviation value would also have to be specified, and such values can be very difficult to obtain. It is clear that other distributions can be used if the appropriate data is available.
During the simulation, the maximum negative NPV, $NPV_{\text{min}}$, during the timeline of the project is recorded. Because no initial inflow of cash is assumed the NPV must initially decline into negative territory and then increase if and when a minimum of one drug is approved for marketing. As shown in Figure 4.1 this cash flow will have a maximum negative value during the simulated lifetime of the project. In reality, a company will have limited cash resources available to devote to the development of a portfolio of drugs at any one time. The cash flow values imposed here are meant to represent the possible cash limitations of a company when funding portfolio development. If this negative cash flow is breached then clearly the company is stretched beyond the limit of finance it had intended to use.

![Figure 4.1. Example of NPV over time for the development of a portfolio of biopharmaceutical drugs.](image)

In order to locate strategies which either do not breach this limit or have low probabilities of doing so, a straightforward penalty method is applied to the set of simulation results produced from the Monte Carlo simulation for each strategy. For each Monte Carlo trial on a strategy, if any value in the projected cash flow over the lifetime of the portfolio’s development is less than $NPV_{\text{min}}$ then, if positive, the terminal NPV indicated by that particular Monte Carlo trial is set to 0. This has the impact of reducing $p(NPV>0)$. Ultimately for strategies that are unable to meet a given cash flow restriction for any particular Monte Carlo trial $p(NPV>0)$ will equal 0 with use of this penalty method.

**4.3. RESULTS AND DISCUSSION**

The results are now presented and discussed.
4.3.1. FIVE DRUG PORTFOLIO

Figure 4.2 shows the final results for a five drug portfolio subject to the various constraints considered. It can immediately be seen that a negative relationship between mean positive NPV and \( p(\text{NPV}>0) \) exists for all constraints. It is also noticeable that although individually each drug has a probability of reaching the marketing phase of between 0.14 and 0.25, depending on the development strategy, the entire portfolio has a \( p(\text{NPV}>0) \) of up to 0.76, over 5 times the lower bound of this range. This relationship highlights a challenge to the drug developer that here appears to be a problem of making decisions when the objectives that must be optimised are also conflicting. It can also be seen in each case that for a given \( p(\text{NPV}>0) \) value the more restrictive the cash flow constraint the lower the mean positive NPV. Similarly, for a given target in mean positive NPV these constraints reduce the probability of achieving this profit. As might be expected, this is indicative that the decision maker in this case needs to consider that by having more financial resources available for the drug development process it is possible to enhance the profitability of the venture. This also suggests that resources critical to the generation of profit are compromised upon the introduction of such constraints. It is also observed in Figure 4.2 that the decline in profitability for a given value of \( p(\text{NPV}>0) \) is most pronounced between the cash flow constraints of \(-$100\text{MM}\) and \(-$75\text{MM}\). The large drop between such a relatively small margin suggests the loss of availability of a key profit-generating resource. Interestingly, the maximum value of \( p(\text{NPV}>0) \) is not reduced by the presence of even the most severe constraints on cash flow investigated. The analysis will proceed by analysing the most probable constituents of strategies for the unconstrained example alongside the \(-$200\text{MM}\) and \(-$100\text{MM}\) constraints. Analysis of the \(-$75\text{MM}\) constraint is not taken further here as it is unrealistic to develop a portfolio of five drugs on such a low level of financing.
Figure 4.2. Mean positive NPV versus p(NPV>0) for a five-drug portfolio under the following constraint levels: unconstrained (-), -$200MM (•), -$100MM(◆), -$75MM (■).

4.3.1.1. THIRD PARTY STRATEGIES

When analysing the third party strategies in the final generation for each constraint (Table 4.1, Table 4.2, and Table 4.3) it is clear that the general trend in third party strategy is to develop and manufacture all drugs in-house under the presence of no constraints and then move towards the increased involvement of partners as constraints become more severe. This is understandable from the viewpoint that increasing the severity of cash flow constraints makes it imperative that more cost-effective strategies be formulated and in the problem formulation used here partners offer the most cost-effective route for drug development. With an unlimited cash flow the dominant strategy is a complete in-house approach however it is not the only possibility amongst the optimised strategies, demonstrating that for the problem considered there are alternative approaches that involve third parties in some way to obtain equivalent and nondominated results. If the complete in-house approach is supported by appropriate strategies for portfolio structure and scheduling strategies that are parsimonious with spending so as to manage the impact of risk during drug development phases then the revenue stream does not have to be compromised. For the -$200MM constraint it can be seen that partners are chosen to be included for the development of either the first drug only or for both the first and second drug. The involvement of the partner is then kept to commercial manufacturing for successive drugs while contractors are used to assist with clinical development and manufacturing. In general, in-house activities appear to be initially constrained to
clinical development for early drugs and then become involved in clinical manufacturing for successive drugs. Also for this constraint it is observed that for the vast majority of strategies presented a third party is selected for commercial manufacturing, indicating that this level of constraint makes in-house commercial manufacturing to be economically unattractive. Finally, the use of partners appears to be the most extensive for the -$100MM constraint. Here partners are used for the majority of activities in clusters 1 and 2. In cluster 3 partners are used more sparingly, with contractors mostly being used for clinical and commercial manufacturing. Consistent with the -$200MM example, commercial manufacturing is also mostly completed by a third party. The above examples clearly show that third parties are an important resource in managing the risk and impact of failure.
Table 4.1. Most probable strategies for a five drug portfolio with no constraints.

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
<th>Drug 4</th>
<th>Drug 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_i</td>
<td>C_{i1}</td>
<td>C_{i2}</td>
<td>C_{i3}</td>
<td>C_{i4}</td>
<td>C_{i5}</td>
</tr>
<tr>
<td>Drug 1</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 2</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 3</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 4</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 5</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 2</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
<th>Drug 4</th>
<th>Drug 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_i</td>
<td>C_{i1}</td>
<td>C_{i2}</td>
<td>C_{i3}</td>
<td>C_{i4}</td>
<td>C_{i5}</td>
</tr>
<tr>
<td>Drug 1</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 2</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 3</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 4</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 5</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 3</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
<th>Drug 4</th>
<th>Drug 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_i</td>
<td>C_{i1}</td>
<td>C_{i2}</td>
<td>C_{i3}</td>
<td>C_{i4}</td>
<td>C_{i5}</td>
</tr>
<tr>
<td>Drug 1</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 2</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 3</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 4</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
<tr>
<td>Drug 5</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
<td>I'</td>
</tr>
</tbody>
</table>

Note: 'I' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, p - the probability of the corresponding strategy being present. C_{i1} through C_{i5} refer to preclinical and clinical trials. C_{i6} through C_{i9} refer to manufacturing for clinical phases and market.
**Table 4.2.** Most probable strategies for a five drug portfolio with a constraint of $400MM.

<table>
<thead>
<tr>
<th>Drug</th>
<th>$C_{v1}$</th>
<th>$C_{v2}$</th>
<th>$C_{l3}$</th>
<th>$C_{l4}$</th>
<th>$C_{l5}$</th>
<th>$C_{l6}$</th>
<th>$C_{l7}$</th>
<th>$C_{l8}$</th>
<th>$C_{l9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 1</td>
<td>'C'</td>
<td>'T'</td>
<td>'P'</td>
<td>'T'</td>
<td>'P'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.51</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.39</td>
</tr>
<tr>
<td>Drug 3</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.39</td>
</tr>
<tr>
<td>Drug 4</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.29</td>
</tr>
<tr>
<td>Drug 5</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Cluster 1

<table>
<thead>
<tr>
<th>Drug</th>
<th>$C_{v1}$</th>
<th>$C_{v2}$</th>
<th>$C_{l3}$</th>
<th>$C_{l4}$</th>
<th>$C_{l5}$</th>
<th>$C_{l6}$</th>
<th>$C_{l7}$</th>
<th>$C_{l8}$</th>
<th>$C_{l9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 1</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.99</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.66</td>
</tr>
<tr>
<td>Drug 3</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.60</td>
</tr>
<tr>
<td>Drug 4</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.28</td>
</tr>
<tr>
<td>Drug 5</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Cluster 2

<table>
<thead>
<tr>
<th>Drug</th>
<th>$C_{v1}$</th>
<th>$C_{v2}$</th>
<th>$C_{l3}$</th>
<th>$C_{l4}$</th>
<th>$C_{l5}$</th>
<th>$C_{l6}$</th>
<th>$C_{l7}$</th>
<th>$C_{l8}$</th>
<th>$C_{l9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 1</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.99</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.69</td>
</tr>
<tr>
<td>Drug 3</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.56</td>
</tr>
<tr>
<td>Drug 4</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.43</td>
</tr>
<tr>
<td>Drug 5</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Cluster 3

<table>
<thead>
<tr>
<th>Drug</th>
<th>$C_{v1}$</th>
<th>$C_{v2}$</th>
<th>$C_{l3}$</th>
<th>$C_{l4}$</th>
<th>$C_{l5}$</th>
<th>$C_{l6}$</th>
<th>$C_{l7}$</th>
<th>$C_{l8}$</th>
<th>$C_{l9}$</th>
<th>$p$</th>
</tr>
</thead>
</table>

Note: 'T' – in-house activity, 'C' – outsourced activity, 'P' – partnered activity, $p$ – the probability of the corresponding strategy being present. $C_{l1}$ through $C_{l5}$ refer to preclinical and clinical trials. $C_{l6}$ through $C_{l9}$ refer to manufacturing for clinical phases and market.
Table 4.3. Most probable strategies for a five drug portfolio with a constraint of $250M.

<table>
<thead>
<tr>
<th></th>
<th>$D_1$</th>
<th>$C_{i1}$</th>
<th>$C_{i2}$</th>
<th>$C_{i3}$</th>
<th>$C_{i4}$</th>
<th>$C_{i5}$</th>
<th>$C_{i6}$</th>
<th>$C_{i7}$</th>
<th>$C_{i8}$</th>
<th>$C_{i9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug 1</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.98</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug 1</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>1.00</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug 1</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.99</td>
</tr>
<tr>
<td>Drug 2</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

Note: 'T' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, $p$ - the probability of the corresponding strategy being present. $C_{i1}$ through $C_{i5}$ refer to preclinical and clinical trials. $C_{i6}$ through $C_{i9}$ refer to manufacturing for clinical phases and market.
4.3.1.2. Drug Selection Strategies

The probable portfolios that are selected in accordance with the constraints can be seen in Table 4.4, with more detailed results shown in Table A.1 in Appendix A. The performance in the objective space has been split into three clusters in order to facilitate the analysis of any trends or patterns that may exist amongst optimised solutions. The first cluster contains strategies that generate high mean positive NPV values and low p(NPV>0) values. The second cluster consists of strategies that exhibit intermediate values of these metrics. The third cluster contains strategies that perform highly in p(NPV>0) but have low mean positive NPV values. For the unconstrained case, the most probable portfolios for clusters 1 and 3 develop two medium demand drugs and three high demand drugs, and for cluster 2 this portfolio is one low demand drug, one medium demand drug and three high demand drugs. Even given the unlimited cash flow resources the dominant drug development strategy is not to develop all four high demand drugs, indicating that such an approach in this case offers a relatively unfavourable balancing of the impact of failure with attainable financial rewards. It should be noted that all three portfolios in the unconstrained case develop three high demand drugs. Under the -$200MM cash flow constraint, as may be expected, the total demand sought to be developed by the most probable portfolio in each cluster under this level of constraint is less than that sought for development in each corresponding cluster in the unconstrained example. The most probable portfolio in cluster 1 develops one low demand drug, two medium demand drugs, and two high demand drugs. In cluster 2 this is one low demand drug, three medium demand drugs and one high demand drug. For cluster 3 this portfolio consists of three medium demand drugs and two high demand drugs. Under the -$100MM constraint the observed portfolios develop for an even smaller demand than under the -$200MM cash flow constraint. For cluster 1 this portfolio is comprised of two low demand drugs, one medium demand drug, and two high demand drugs. In cluster 2 this is two low demand drugs, two medium demand drugs, and one high demand drug. In cluster 3 this portfolio is one low demand drug, three medium demand drugs, and one high demand drug. More generally, the constraints force successful strategies to form portfolios that cater for a lower demand for all clusters. This is one contributor to why the mean expected NPV decreases with increasing magnitudes of cash flow.
constraint. The level of reduction is generally equal across clusters 1, 2, and cluster 3. Across all clusters it can be seen that in most cases a mixture of drugs from all groups are chosen. As might be expected, drugs from group 3 are most commonly seen in cluster 3 for each level of constraint, as group 3 drugs have the highest probability of success. Additionally, group 3 drugs are scarcely seen in clusters 1 and 2.

Table 4.4. Characteristics of the most probable drug selection and timing strategies for a 5 drug portfolio subject to cash flow constraints.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Drug group</th>
<th>Timing</th>
<th>Characteristic</th>
<th>Demand</th>
<th>Unconstrained</th>
<th>-$200MM</th>
<th>-$100MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 1 2 3 3</td>
<td>ID F III M -</td>
<td>1 2 3 4 5</td>
<td>M M H H H</td>
<td>M M H H L</td>
<td>L H M H L</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 2 1 2 2</td>
<td>P P M F -</td>
<td>1 2 3 4 5</td>
<td>M L H H H</td>
<td>M M H L M</td>
<td>M L M H L</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2 1 3 1 3</td>
<td>ID F II I -</td>
<td>1 2 3 4 5</td>
<td>M L H H H</td>
<td>M M H M L</td>
<td>M L L H M</td>
<td></td>
</tr>
</tbody>
</table>

Note: Demand – relative level of annual demand, AGR – annual growth rate, Drug group – the indication groups most likely to be selected, and Timing – timing strategy for the portfolio. M, L, and H refer to low, medium, and high respectively. 1 through 5 in the header refer to the first through fifth drugs to be developed. 1, 2, and 3 in the drug group rows refer to drug groups 1, 2, and 3. In the timing rows: ID – target identification, P – preclinical trials, I – phase I clinical trials, II – phase II clinical trials, III – phase III clinical trials, F – FDA review, and M – marketing stage. ‘-‘ means that the datum is irrelevant for that particular combination of referencing items.

4.3.1.3. Timing Strategies

The most probable timing strategies can be viewed according to cash flow constraint level and cluster is as shown in Table 4.4. The popular theme exhibited by strategies generated under no cash flow constraints appears to be to develop an initial collection of drugs in close succession or even together and then to develop the remaining drugs with longer intervals between them. As seen in cluster 2 with this level of constraint,
this initial collection can be as many as three drugs. With an unconstrained cash flow
the most probable timing strategies for clusters 1, 2, and 3 take 35 years, 31 years, and
28 years for completion of portfolio development and marketing activities, respectively. Specifically, research and development activities on the portfolio are
expected to be completed 10 years prior to each of these completion times, which is
true for all figures of this type reported in the results section unless otherwise stated.
This indicates a trend of shorter completion times with increasing probability of
success. In cluster 1 two drugs are developed together with the remaining drugs being
developed almost one at a time due to the lengthy intervals imposed by the strategy.
In cluster 2 three drugs are developed within a year of each other, while the remaining
drugs are developed with at least a nine year interval between them. In cluster 3, the
first two drugs are developed together, the third drug is developed after a nine year
interval, and then the two remaining drugs are developed with medium length
intervals of four years in between them. Under the -$200MM cash flow constraint, in
all clusters the portfolio development times become longer. Clusters 1, 2, and 3 take
38 years, 38 years, and 40 years for the completion of portfolio development and
marketing activities, respectively. The trend exhibited with the unconstrained
example no longer exists here and the development times across the clusters have
become almost equal. If any trend exists then it is that the development times are
longer towards the higher values of p(NPV>0). The theme of developing at least two
drugs in close succession is also not maintained. As shown in Table 4.4 in all clusters
the most probable strategy is to wait until drug 1 is expected to begin the FDA review
before developing drug 2. Interestingly, the development times for drugs 2, 3, and 4
are brought into closer succession for clusters 1 and 2. For cluster 3 this is not the
case and instead an alternation of waiting times is seen between waiting until the FDA
review and then until phase III clinical trials for the next drug. Under the -$100MM
cash flow constraint the total completion time for development and commercialisation
is 32, 32, and 37 years, respectively. It is possible that the intensive involvement of a
partner, as shown earlier, allows for this to happen by serving as an effective scheme
for managing the impact of failure. The intervals after drugs 2 and 3 are shorter for
clusters 1 and 2 when compared to the same clusters for the -$200MM constraint. It
is thought that in clusters 1 and 2 any cash flow arising from revenue of the first drug
will fuel the development of some subsequent drugs allowing for shorter intervals to
be applied up to a point. In cluster 3, the intervals have stayed identical to those
correspondent with the less restrictive constraint considered and this is thought to be the case because this cluster takes on portfolios that cater for a much larger demand than the other two clusters. The most observable trend across the constraints is that the development time between drugs 2, 3, and 4 becomes consistently shorter in clusters 1 and 2. One way of mitigating the impact of failure is to lengthen the time interval between the developments of subsequent drugs such as to allow for the possibility of previous drugs to reach market and provide a cash flow to fund future projects.

4.3.2 Three Drug Portfolio

The final results for a three drug portfolio can be seen in Figure 4.3 where the effect of constraints on the final population is clear. Again as with the five drug portfolio the impact of constraints reduces the mean positive profit for a given p(NPV>0) value. It is immediately noticeable that the probability of the portfolio yielding a positive NPV is significantly reduced when compared to five drug portfolio. It is also noticeable that the range of mean positive NPV values for the three drug portfolio is effectively the same as for the five drug portfolio suggesting that both portfolios are relying on the same number of drugs reaching market, which is most probably one drug. This is because drug development is uncertain and drugs are more likely to fail than not during this process. Thus with portfolios of decreasing size there is a decreased likelihood that at least one candidate will reach the market place. It can be seen that the frontier of the -$100MM constraint is of closer proximity to the frontier of the unconstrained scenario than to that of the -$75MM constraint. This signifies the loss of some set of critical strategies with the -$75MM constraint that would otherwise contribute significantly to profit generation and the mitigation of unwanted risk in this region. Figure 4.3 also shows that the topology of the frontier for the -$100MM constraint lies similar to that for the unconstrained problem above a mean positive NPV of $1,000MM. There is even a point along the frontier for the -$100MM constraint at approximately p(NPV>0) = 0.3 where it is equal to the unconstrained problem. Below this measure of profit the profile of the frontier produced by the -$100MM constraint inverts demonstrating the largest loss of value and strategies as imposed by the constraint is in this region.
4.3.2.1. Third Party Strategies

As seen in Table 4.5, Table 4.6, and Table 4.7 all strategies take a completely different approach to the strategic involvement of third parties as with those formulated for the five drug portfolio. The most popular strategy amongst these is to assign the majority of development activities for the first drug to a partner and then to develop successive drugs with the involvement of a contractor. In the unconstrained problem, the approach to include partners in this way is phased out towards cluster 3 where in-house and outsourcing strategies dominate. When cash flow constraints do not apply it can be seen that the involvement of third parties is minimal when compared to third party strategies seen when such constraints are applied. With the -$100MM constraint, more partners are included towards the third cluster where a complete partnered approach is the most probable third party strategy. The -$75MM constraint yields strategies that use partners more in clusters 1 and 3 than in cluster 2. The probable approach of intensive use of partners and contractors in this manner highlights the necessity for parsimonious cash flow management with the three drug portfolio when compared to the five drug portfolio. The use of partners during early stages is understandable as in this case they allow for the most economically efficient route to market. The clear trade-off is that a large proportion of the revenue must be shared and it is economically inefficient to continue with this strategy unless forced to do so by cash constraints. Limiting subsequent use of a partner to commercial
manufacturing minimises the proportion of revenue that will have to be paid and it also allows the company to overcome the significant cost of acquiring commercial manufacturing facilities. As this facility is scheduled to be built at the beginning of phase III clinical trials for the drug that it will be manufacturing there is a reasonable likelihood that the drug will fail leaving the company with a sunk capital expense. Clearly, partnering reduces this impact of failure.

Table 4.5. Most probable strategies for a three drug portfolio with no constraints.

<table>
<thead>
<tr>
<th></th>
<th>$D_1$</th>
<th>$C_{i1}$</th>
<th>$C_{i2}$</th>
<th>$C_{i3}$</th>
<th>$C_{i4}$</th>
<th>$C_{i5}$</th>
<th>$C_{i6}$</th>
<th>$C_{i7}$</th>
<th>$C_{i8}$</th>
<th>$C_{i9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Drug 1</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Drug 2</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Drug 3</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'P'</td>
<td>'P'</td>
<td>0.48</td>
</tr>
</tbody>
</table>

|        | Drug 1 | 'T'      | 'T'      | 'T'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 0.85|
|        | Drug 2 | 'T'      | 'T'      | 'T'      | 'C'      | 'C'      | 'C'      | 'C'      | 'C'      | 'C'      | 0.85|
|        | Drug 3 | 'C'      | 'T'      | 'T'      | 'T'      | 'T'      | 'T'      | 'C'      | 'P'      | 'P'      | 0.85|

|        | Drug 1 | 'T'      | 'T'      | 'T'      | 'T'      | 'T'      | 'T'      | 'C'      | 'C'      | 'C'      | 0.52|
|        | Drug 2 | 'C'      | 'C'      | 'T'      | 'C'      | 'C'      | 'T'      | 'C'      | 'P'      | 'P'      | 0.52|
|        | Drug 3 | 'C'      | 'T'      | 'T'      | 'T'      | 'C'      | 'T'      | 'C'      | 'P'      | 'P'      | 0.52|

Note: 'T' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, $p$ - the probability of the corresponding strategy being present. $C_{i1}$ through $C_{i5}$ refer to preclinical and clinical trials. $C_{i6}$ through $C_{i9}$ refer to manufacturing for clinical phases and market.

Table 4.6. Most probable strategies for a three drug portfolio with a -$100MM constraint.

<table>
<thead>
<tr>
<th></th>
<th>$D_1$</th>
<th>$C_{i1}$</th>
<th>$C_{i2}$</th>
<th>$C_{i3}$</th>
<th>$C_{i4}$</th>
<th>$C_{i5}$</th>
<th>$C_{i6}$</th>
<th>$C_{i7}$</th>
<th>$C_{i8}$</th>
<th>$C_{i9}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Drug 1</td>
<td>'T'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>'P'</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Drug 2</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'T'</td>
<td>'P'</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Drug 3</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>'C'</td>
<td>'T'</td>
<td>'T'</td>
<td>0.55</td>
</tr>
</tbody>
</table>

|        | Drug 1 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 1.00|
|        | Drug 2 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 0.81|
|        | Drug 3 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 0.52|

|        | Drug 1 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 1.00|
|        | Drug 2 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 1.00|
|        | Drug 3 | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 'P'      | 1.00|

Note: 'T' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, $p$ - the probability of the corresponding strategy being present. $C_{i1}$ through $C_{i5}$ refer to preclinical and clinical trials. $C_{i6}$ through $C_{i9}$ refer to manufacturing for clinical phases and market.
Table 4.7. Most probable strategies for a three drug portfolio with a -$75MM constraint.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
<th>C_1</th>
<th>C_2</th>
<th>C_3</th>
<th>C_4</th>
<th>C_5</th>
<th>C_6</th>
<th>C_7</th>
<th>C_8</th>
<th>C_9</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>T 'P' 'P' 'P' 'P' 'P' 'P' 'P' 'P' 0.422</td>
<td>T 'P' 'P' 'P' 'P' 'P' 'P' 'P' 'P' 0.281</td>
<td>T 'C' 'T' 'T' 'T' 'T' 'T' 'T' 'T' 0.281</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>T 'C' 'P' 'P' 'P' 'P' 'P' 'P' 'P' 0.198</td>
<td>T 'C' 'P' 'P' 'P' 'P' 'P' 'P' 'P' 0.228</td>
<td>T 'C' 'T' 'T' 'T' 'T' 'T' 'T' 'T' 0.216</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>T 'P' 'P' 'P' 'P' 'P' 'P' 'P' 'P' 0.409</td>
<td>T 'C' 'T' 'T' 'T' 'T' 'T' 'T' 'T' 0.309</td>
<td>T 'T' 'T' 'T' 'T' 'T' 'T' 'T' 'T' 0.271</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 'T' - in-house activity, 'C' - outsourced activity, 'P' - partnered activity, p - the probability of the corresponding strategy being present. C_1 through C_9 refer to preclinical and clinical trials. C_10 through C_19 refer to manufacturing for clinical phases and market.

4.3.2.2. DRUG SELECTION STRATEGIES

The drug selection strategies for the most probable portfolio for each cluster in the optimised population of strategies can be viewed in Table 4.8. More detailed results can be viewed in Table A.2 in Appendix A. Across the entire range of constraints and clusters there are no discernable trends with the demand of drugs that are consistent over all constraints. All portfolios select drug candidates that have a mixture of demands, rather than what might be thought to be an aggressive profit-focussed strategy such as selecting all high demand drugs. All portfolios contain at least one high demand drug and no portfolio opts to develop more than one low demand drug. High demand drugs are developed second or third by the majority of strategies in Table 4.8. Additionally, relatively few strategies in Table 4.8 choose drugs with high annual growth rates.
Table 4.8. Probabilities of drug selection for a three drug portfolio.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Unconstrained</th>
<th>-$100MM</th>
<th>-$75MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand</td>
<td>M H L</td>
<td>M M H</td>
<td>M H L</td>
</tr>
<tr>
<td>AGR</td>
<td>M L H</td>
<td>M L L</td>
<td>M M M</td>
</tr>
<tr>
<td>Drug group</td>
<td>2 1 3</td>
<td>2 1 1</td>
<td>1 2 1</td>
</tr>
<tr>
<td>Timing</td>
<td>M M -</td>
<td>M M -</td>
<td>M M -</td>
</tr>
</tbody>
</table>

Cluster 1

| Demand         | M H H          | M M H   | M H H  |
| AGR            | M L L          | M L L   | M L H  |
| Drug group     | 2 2 1          | 2 2 3   | 1 2 1  |
| Timing         | M M -          | M M -   | M M -  |

Cluster 2

| Demand         | M H H          | M L H   | H L M  |
| AGR            | L M H          | L M L   | L H L  |
| Drug group     | 2 2 2          | 2 1 3   | 1 3 1  |
| Timing         | M M -          | M M -   | F M -  |

Cluster 3

Note: Demand – relative level of annual demand, AGR – annual growth rate, Drug group – the indication groups most likely to be selected, and Timing – timing strategy for the portfolio. M, L, and H refer to low, medium, and high respectively. 1 through 5 in the header refer to the first through fifth drugs to be developed. 1, 2, and 3 in the drug group rows refer to drug groups 1, 2, and 3. In the timing rows: ID – target identification, P – preclinical trials, I – phase I clinical trials, II – phase II clinical trials, III – phase III clinical trials, F – FDA review, and M – marketing stage.

4.3.2.3. TIMING STRATEGIES

Under almost all constraints for all clusters the optimised strategies opt to extend the development of the portfolio to as long as possible (Table 4.8). In clusters 1 and 2 there is an increase in the probability by which these timing strategies are selected. It is clear that the constraints on the portfolio limit the number of projects that can be undertaken simultaneously. What might be unintuitive is that even with unlimited cash flow constraints this remains the most probable strategy and alternative timing strategies to this follow a probability distribution that becomes less probable for selecting more time prudent alternatives. This is very much different to the approach on timing seen for the five drug portfolio where usually two or three drugs are often either developed together or in close succession with the remaining drugs staggered into longer phases. Because there is a significantly greater probability that the entire three drug portfolio will fail to reach market and produce a profit, it becomes less
appropriate to develop multiple drugs together. In this case it is most appropriate to completely disaggregate the cash flows of drug development projects.

4.4. CONCLUSIONS

The technique presented has been used to explore the difficult task of identifying optimal strategic approaches to the development of a portfolio of drugs. Analyses using a five drug portfolio and a three drug portfolio have been presented. Additionally, cash flow constraints on these portfolios have been applied. It has been seen in the cases investigated here that mean positive NPV and $p(NPV>0)$ are conflicting measures however both are desirable to the decision maker. The introduction of cash flow constraints can lead to a reduction in the expected rewards or probability of success of strategy. Cash flow constraints also directly influence third party strategies. The reduction in portfolio size from five drugs to three drugs produces the same mean positive NPV range but results in a significant reduction in the probability that the portfolio returns a profit. The timing and third party strategies for each size of portfolio are very different showing that the size of the portfolio has a specific impact on how it should be developed and commercialised regardless of how much capital is available. This signals to the decision maker that it may not be appropriate to apply similar development strategies to portfolios of different sizes. It is important to remember that the results are dependent on the nature of the scenario that is to be investigated. In this case, if drugs that were already partially through the clinical trial process had been considered then it is possible that a very different set of results may have been generated by the framework. On the issue of comparability, it is difficult to compare the results here to that of other works because of the individual properties of the case study. Also, because each strategy within the set of optimised strategies is comprised of decisions that can add value collectively rather than just individually it is difficult to compare the elements that are common with similar works. Finally, the reader should be aware that this is a computationally intensive procedure which through the use of a Pentium IV processor can take approximately 15 hours to execute without the use of parallel processing techniques. Hence depending on available computational resources extending model variables such as the number of Monte Carlo trials or the size of the portfolio can come at the significant expense of time. Specifically, it is important to note that the
computational limits of this framework are dependent on a few factors: the processing capacity of the computational system under use, the programming language that it is executed in, multi-threaded design of the program, and the allowable execution time.
CHAPTER 5

Economic Correlations for Whole Monoclonal Antibody Manufacture and Antibody Fragments Manufacture

5.1. INTRODUCTION

In chapters 2-4 methods for rationalising between strategic options, optimising superstructures of strategies, and analyses for determining the impact of factors affecting these processes have been investigated. These studies have all required the development of mathematical models for simulating the capital expenditure and the operating cost incurred in constructing biopharmaceutical manufacturing facilities. Such models require computationally expensive evaluation procedures for determining the required economic metrics. The level of modelling employed permits a rich level of detail to be provided and is particularly useful in the event that a discerning detail-orientated decision maker wishes to access such information. It is intuitive that there are situations where computational efficiency is also of great interest. Such situations might include when using the optimisation framework presented in chapter 3 which must perform numerous Monte Carlo samples for several thousand strategies. Also from a more practical perspective, decision makers in the biopharmaceutical industry could benefit from access to simplified formulae for calculating key economic metrics of interest, such as fixed capital investment and the cost of goods sold per gram (COGS/g). In either of these cases, it is obvious that a less computationally intensive estimation method that is able to provide an acceptable level of accuracy would be of considerable use. This is the focus of this chapter.

Specifically, the scope of investigation here is focused on the development of formulae that approximate the economic outputs of detailed mathematical biopharmaceutical manufacturing models. A common straightforward approach has previously been used in industry to estimate the cost of individual items of bioprocessing equipment, with the following relationship between size and cost employed otherwise known as the $R$ factor method Remer and Idrovo (1991):
This technique inherently assumes a geometric relationship between variables. The exponent $R$ for this correlation can be found through linear regression of size and cost data on log-log axes, where $R$ is the gradient of the line of best fit. Such methods commonly set this value to 0.6 although this is not ubiquitous for all equipment. This method was first applied by Williams (Williams, 1947) for equipment costing and later by Chilton (1950) for plant costing. As the biopharmaceutical industry has grown, the value of the exponent $R$ has been investigated for commonly used equipment often used in this industry (Remer and Idrovo, 1991; Remer and Chai, 1990). Remer and Idrovo (1991) have conducted empirical evaluations of $R$ for a large range of upstream and downstream bioprocessing equipment. Their results indicating that $R$ varies between 0.37 and 1.16 and that the average value of $R$ for the dataset used in their investigation was 0.63. More recently $R = 0.6$ has been used in literature for cost models in the optimisation of the design of biopharmaceutical plants (for example see Dietz et al., 2007; Dietz, et al., 2006). The above indicates that geometric expression of the relationship between the size and cost of bioprocessing equipment has been accepted in literature. Also the above highlights that this technique has been useful for approximating the cost of individual items of equipment, as seen in Remer and Idrovo (1991), as well as for building more intricate cost models of biopharmaceutical processes, as seen in Dietz et al.(2006 and 2007). Some limitations of this technique have been noted by Remer and Idrovo (1991). Some types of bioprocess equipment such as high performance liquid chromatography are difficult to scale up using the $R$ factor method. As there must be a referencing value for size and cost to use the technique Remer and Idrovo (1991) note that there is a range of validity beyond which use of this method becomes inappropriate.

The approach taken by this work is to extend the concept of geometric relationships used to approximate the cost of individual items of equipment, and also to investigate if a straightforward geometric expression can be used to estimate the key economics of an entire facility. When determining cost estimating equations for equipment, the
R factor technique will not be used in place of regression analysis techniques where appropriate. This is because regression analysis techniques can characterise an entire dataset without the need for base values to be used in the equation itself, thus potential issues such as ranges of validity are avoided. This requires the development of a detailed mathematical model of specific biopharmaceutical manufacturing processes so that costs are accurately determined. The next step draws computationally straightforward relationships between key input factors and the broad economic metrics of interest to a decision maker in a biopharmaceutical setting. Such metrics are typically the fixed capital investment required to build a biomanufacturing facility as well as its operating cost. The approximation method used to achieve this here is multiple regression analysis (Kitchens, 1998) which is used to learn nonlinear relationships between the inputs and outputs of the model. Two types of manufacturing process will be considered, that is, the manufacture of whole monoclonal antibodies using mammalian cells, and the manufacture of antibody fragments using *E.coli*. The following sections will entail:

- Description of each manufacturing process with regard to the unit operations that comprise them.
- Mathematical modelling of the relevant unit operations
- Analysis of model outputs over a range of industrially relevant inputs.
- Development of the approximation formulae.
- Analysis on the error when predicting the outputs of the fully detailed mathematical models.

### 5.2. Method

The method for the modelling, simulation, and approximation of the economics of whole monoclonal antibodies and antibody fragments manufacture is now presented.

#### 5.2.1 Process Overviews

The production process for the manufacture of whole monoclonal antibodies is exhibited in Figure 5.1. The choice of process operations for this product was largely influenced by similar processes of mammalian cell culture-derived monoclonal antibody drugs such as *Herceptin™*, *Zenapax™*, *Rituxan™*, *Synagis™*, and
Remicade™ (Farid, 2001). Also, in Figure 5.2 the process for manufacturing antibody fragments is presented which is based on the manufacturing procedure for Celltech's E.coli based production of PEG-conjugated antibody fragments, CDP870 (Farid, 2001). Accordingly, the mathematical models constructed to replicate these processes include the unit operations required for production and processing, as well as the ancillary tasks, raw materials, utilities and labour. There are four inputs to this model: the annual level of market demand that must be met by the facility, the fermentation titre that will be achieved, the product yield after downstream processing, and the probability of achieving successful batch fermentation. These inputs were used to calculate the masses and volumes of products and wastes that will be generated during the manufacturing process. These were subsequently used to calculate the appropriate number and size of equipment, alongside the raw materials and utilities utilised. The data used for equipment prices was collected from a variety of commercial sources. Correlations between the price and size of each item of equipment or raw material were derived and the model uses the correlation-based equations to evaluate the economics of the required manufacturing process. In cases where only fixed denominations of size were available the model references the cost for the appropriate size. These models are implemented as spreadsheet architectures in MS Excel.

5.2.2 Unit Operations

Key inputs and outputs to the mathematical models for calculation of equipment costs and sizing are hereon described. Equipment and mass balance equations are based on work by Farid (2001), with full details provided in Appendix B.
Figure 5.1. The whole monoclonal antibody biopharmaceutical manufacturing process model. Abbreviations: DSP - downstream processing, QC/QA – quality control / quality assurance, FCI – fixed capital investment, COGS – cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. DSP yield is the overall yield after all downstream processing steps have been completed. Batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.
Figure 5.2. The antibody fragments biopharmaceutical manufacturing process model. Abbreviations: DSP - downstream processing, QC/QA - quality control / quality assurance, FCI - fixed capital investment, COGS - cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. DSP yield is the overall yield after all downstream processing steps have been completed. Batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.
Table 5.1. Key inputs and outputs for key operations.

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td>Annual demand</td>
<td>Fixed capital investment</td>
</tr>
<tr>
<td></td>
<td>Titre (g/L)</td>
<td>COGS/g</td>
</tr>
<tr>
<td></td>
<td>Whole process yield</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Batch success probability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Batches per year</td>
<td></td>
</tr>
<tr>
<td>Inoculum</td>
<td>Process cycle time</td>
<td>Outflow volume</td>
</tr>
<tr>
<td></td>
<td>Inoculum volume required</td>
<td></td>
</tr>
<tr>
<td>Fermentation</td>
<td>Process cycle time</td>
<td>Outflow stream composition</td>
</tr>
<tr>
<td></td>
<td>Production volume required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stoichiometry coefficients</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inflow stream composition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Media required</td>
<td></td>
</tr>
<tr>
<td>Filtration</td>
<td>Inflow stream composition</td>
<td>Outflow stream composition</td>
</tr>
<tr>
<td></td>
<td>Average flux (L/m²/h)</td>
<td>Retentate tank size</td>
</tr>
<tr>
<td></td>
<td>Concentration factor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rejection coefficient</td>
<td></td>
</tr>
<tr>
<td>Chromatography</td>
<td>Inflow stream composition</td>
<td>Outflow stream composition</td>
</tr>
<tr>
<td></td>
<td>Chromatography steps</td>
<td>Buffer holding tank sizes</td>
</tr>
<tr>
<td></td>
<td>Column volumes of requisite buffers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resin binding capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resin reusability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flow rate</td>
<td></td>
</tr>
<tr>
<td>Centrifugation</td>
<td>Inflow stream composition</td>
<td>Outflow stream composition</td>
</tr>
<tr>
<td></td>
<td>Flow rate</td>
<td>Supernatant holding tank size</td>
</tr>
<tr>
<td></td>
<td>Solid carry-over level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cycle time</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Equipment numbers, equipment costs, raw materials costs, and operating costs are outputs for all operations. All operations require details of the inflow stream composition as an input.*
Table 5.1 displays the principle inputs and outputs to the simulation models as well as to individual unit operations. The overarching inputs to the models are the annual demand for the facility, titre of product to be achieved at fermentation, whole process yield, batch success probability, and the number of batches to be run annually. It is important to note that the number of batches to be run annually is more dependent on the nature of the manufacturing process rather than on decisions made for design of the manufacturing facility. This is mainly because of the time required for the host expression system to achieve the desired titre of crude product. In the mammalian-derived whole monoclonal antibody process it is assumed that 20 batches per year are produced, whereas for the bacterial production of antibody fragments this number is 50. These numbers are based on the time taken to produce the crude product and cascaded scheduling of the manufacturing process such that fermentation can be recommenced immediately after the bioreactor has been sterilised. Because no operations in the downstream process for either manufacturing procedure are longer than the fermentation stage no bottlenecks in the downstream operations are created in the model by use of this design. Inputs to unit operations beyond the four main inputs are derived from within each model and do not have to be provided by the user once the model has been programmed. For example, inputs such as the inflow stream composition for a particular unit operation are taken from the outflow stream of its preceding operation. This data can be subsequently used to determine the number of equipment units required as well as their sizes using mathematical relationships. Raw materials and operating costs can also be determined by calculating their usage and referencing a database containing relevant costs. Ultimately all equipment numbers, equipment costs, raw material costs, and operating costs are calculated from mathematically relating the mass composition of the input stream to the requisite operational protocols of individual equipment units.
Table 5.2. Rules for adding additional equipment units.

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Rule for adding new equipment units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoculum flask</td>
<td>Processing volume &gt; 0.5L</td>
</tr>
<tr>
<td>Fermentation</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>Ultrafiltration unit</td>
<td>Processing time &gt; 24h</td>
</tr>
<tr>
<td>Affinity chromatography column</td>
<td>Column diameter &gt; 0.45m</td>
</tr>
<tr>
<td>Virus inactivation tank</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>Cation exchange chromatography column</td>
<td>Column diameter &gt; 0.45m</td>
</tr>
<tr>
<td>Virus nanofiltration unit</td>
<td>Processing time &gt; 24h</td>
</tr>
<tr>
<td>Diafiltration unit</td>
<td>Processing time &gt; 24h</td>
</tr>
<tr>
<td>Anion exchange chromatography column</td>
<td>Column diameter &gt; 0.45m</td>
</tr>
<tr>
<td>Dead-end filtration</td>
<td>Column diameter &gt; 0.45m</td>
</tr>
<tr>
<td>Periplasmic extraction tank</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>Tank for pH and conductivity adjustment</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>Dilution tank</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>PEGylation tank</td>
<td>Processing volume &gt; 20,000L</td>
</tr>
<tr>
<td>Centrifugation</td>
<td>Processing time &gt; 24h</td>
</tr>
</tbody>
</table>

Note: Above equipment dimensions represent maximum dimensions available at the time of study.

Table 5.2 displays the protocols for adding new equipment units across both manufacturing processes. These are based on maximum sizes of commercially available equipment and processing times that would not create bottlenecks in the manufacturing process. Realistic denominations of equipment sizes were used, which included sizes of marketed equipment where available. By ensuring that processing time did not introduce bottlenecks scalability of the model was supported in determining appropriately equipped facilities that could operate in line with given time constraints. It should be noted that when a new fermenter is added its outflow stream is merged with that of other fermenters, rather than creating a new downstream processing train. This is true of any other equipment unit. Also, equipment sizes are calculated such that multiple units of the same equipment are all sized equally.
5.2.3 COST OUTPUTS

The output costs of the models are now presented. The fixed investment cost is calculated by linear correlation to the total equipment cost using the Lang Factor method (Lang, 1948):

$$FCI = E_{EQP} \sum L_{f,i}$$  (5.2)

Where $FCI$ is the fixed capital investment required to build the facility, $E_{EQP}$ is the total equipment cost of the facility, and $L_{f,i}$ is the value of the $i^{th}$ factor of the relevant Lang factors. The breakdown of the Lang factors (Lang, 1948; Novais, Titchener-Hooker, and Hoare, 2001) are shown in Table 5.3. The model used for the cost of goods is seen in Table 5.4.

Table 5.3. Lang factor constituents and values.

<table>
<thead>
<tr>
<th>Description</th>
<th>$f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equipment and utilities</td>
<td>1.00</td>
</tr>
<tr>
<td>Pipework and installation</td>
<td>0.90</td>
</tr>
<tr>
<td>Process control</td>
<td>0.37</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>0.60</td>
</tr>
<tr>
<td>Electrical power</td>
<td>0.24</td>
</tr>
<tr>
<td>Building works</td>
<td>1.66</td>
</tr>
<tr>
<td>Detail engineering</td>
<td>0.77</td>
</tr>
<tr>
<td>Construction</td>
<td>0.40</td>
</tr>
<tr>
<td>Commissioning</td>
<td>0.07</td>
</tr>
<tr>
<td>Validation</td>
<td>1.06</td>
</tr>
<tr>
<td>Contingency factor</td>
<td>1.15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8.13</strong></td>
</tr>
</tbody>
</table>

Source: Novais et al. (2001).
ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY MANUFACTURE AND
ANTIBODY FRAGMENTS MANUFACTURE

Table 5.4. COGS model breakdown.

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>Description</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct raw materials</td>
<td>f(utilisation)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>Miscellaneous materials</td>
<td>0.5 x direct raw materials</td>
<td>(5.4)</td>
</tr>
<tr>
<td>Direct utilities</td>
<td>f(utilisation)</td>
<td>(5.5)</td>
</tr>
<tr>
<td>Operating labour</td>
<td>f(utilisation)</td>
<td>(5.6)</td>
</tr>
<tr>
<td>Supervisors</td>
<td>0.2 x operating labour</td>
<td>(5.7)</td>
</tr>
<tr>
<td>Quality control and quality assurance</td>
<td>1 x operating labour</td>
<td>(5.8)</td>
</tr>
<tr>
<td>General management</td>
<td>1 x operating labour</td>
<td>(5.9)</td>
</tr>
</tbody>
</table>

| Indirect Cost of Goods | Maintenance 0.1FY           | (5.10)   |
| Local taxes           | 0.02FY                      | (5.11)   |
| Insurance             | 0.01FY                      | (5.12)   |
| Capital charge        | F(1 + rate of interest)^C-1  | (5.13)   |
| General utilities     | Cost per unit area per year x Facility size x Y | (5.14) |

Total Cost of Goods  Direct Cost of Goods + Indirect Cost of Goods (5.15)
Total Cost of Goods per Gram (COGS)  Total Cost of Goods / Annual Production Output (5.16)

Note: Where F is the fixed capital investment, Y is the project duration in years, and C is the capital charge period. Source: Mustafa et al. (2004).

In the model, the facility size was estimated through a correlation determined from data presented in Pavlotsky (2004) and Farid (2007). Pavlotsky (2004) takes a detailed approach to approximating facilities cost and includes more than 100 pharmaceutical projects. The facilities in the study were categorised into three size ranges: 25,000 sq.ft, 50,000 sq.ft and 100,000 sq.ft with the encompassing cost data being adjusted to fit these categories. Farid (2007) includes data on recently built antibody manufacturing facilities that use mammalian cell culture. The size to cost data from these studies are plotted in Figure 5.3 with a geometric relationship being strongly supported.
ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY MANUFACTURE AND ANTIBODY FRAGMENTS MANUFACTURE

Figure 5.3. Facility size versus cost relationship based on data from Pavlotsky (2004) and Farid (2007). The solid line represents the line of best fit as determined by geometric regression. The equation of this line is \( f(x) = 1730x^{0.9136} \), with \( R^2 = 0.9539 \).

### 5.2.4 Simulation Method for Data Acquisition

Estimation of key process economic metrics was investigated by determining relationships via multiple regression analysis to the four main inputs: annual demand \((D)\), fermentation titre \((T)\), whole process yield \((Y)\), and batch success probability \((p_B)\). In order to acquire the data necessary for investigating relationships, different combinations of instances of these inputs were inputted to the model with the outputs recorded. Table 5.5 summarises the range of instances for these inputs. The number of batches produced by the facility per year, \(n_B\), was kept constant at 20 batches/year for mammalian-derived monoclonal antibody production and 50 batches per year for bacterial derived antibody fragment production.

<table>
<thead>
<tr>
<th>Input</th>
<th>Range</th>
<th>Number of instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D)</td>
<td>10kg, 25-500kg at 25kg intervals</td>
<td>21</td>
</tr>
<tr>
<td>(Y)</td>
<td>0.4, 0.5, 0.6, 0.7</td>
<td>4</td>
</tr>
<tr>
<td>(T)</td>
<td>0.25, 0.5, 1.0, 1.5, 2.0 g/L</td>
<td>5</td>
</tr>
<tr>
<td>(p_B)</td>
<td>0.7, 0.8, 0.9, 1.0</td>
<td>4</td>
</tr>
</tbody>
</table>

The ranges in Table 5.5 were based on what is currently achieved in industry plus some reasonable extension that might represent future developments at the time of
study (2004). In terms of titre, the typical maximum value observed in the industry in 2004 was about 1g/L although higher titres of 3-5g/L are now seen in 2008. At the time of study it was sensible to choose values ranging from 0.25g/L to 2g/L (Sommerfeld & Strube, 2005). Resultantly, a set of 1680 input combinations was used to build correlations to the outputs. The outputs measured from the simulation were:

- Fixed capital investment
- Upstream process (USP) equipment cost
- Downstream process (DSP) equipment cost
- COGS/g
- Contribution (%) of material, labour, and indirect costs comprising COGS/g

The reasons for selecting these outputs are now explained. Capital investment and cost of goods related metrics are of clear importance to decision makers in the industry. Determination of the USP and DSP costs as well as the ratio between them can assist in identifying the most influential driver of capital expense. COGS/g indicates the profitability of the manufacturing process and implies its efficiency. The contribution of key components to COGS/g will identify its most influential driver.

5.2.5 Multiple Regression Technique

Multiple regression aims to establish the nature of relationships between a set of independent variables, and a dependent variable (Kitchens, 1998). In this circumstance the independent variables are $D$, $T$, $Y$, and $p_B$. The dependent variable can be any one of the outputs previously mentioned. The technique is able to determine the extent to which a set of independent variables explains the variance in a dependent variable through a significance test. The measure of significance used here is the coefficient of determination, $R^2$. The independent variables are the input variables, which are assumed to have no correlation with each other. The dependent variable is the output metric of interest that is determined by some computation on the independent variables. The multiple regression technique can be used to learn a range of relationships including linear, hyperbolic, geometric, and polynomial relationships. This study focuses on exploring the derivation of geometric relationships between the model inputs and outputs. This is partly because, as previously stated, geometric relationships for estimating economic metrics in relevant settings have received
acceptance in published literature. Geometric relationships as used here will also lead to straightforward formulae. Also, they were generally found here to have the least error in explaining model outputs when compared to linear and hyperbolic relationships. Deriving polynomial relationships is more involved and relies on specifying the maximum exponent value in the polynomial equation. As the order of the polynomial equation increases, more reliability in the estimation of outputs is expected but the trade-off is relationships that are more expansive and less convenient. For this reason polynomial relationships were not considered here.

Geometric relationships between the independent variables and the dependent variable in the multiple regression model take the following form:

\[ \ln(X) = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \beta_3 \ln(x_3) + ... + \beta_n \ln(x_n) + \epsilon \]  

(5.17)

Where \( X \) – a variable dependent on the values of \( x_1 \) through \( x_n \), \( \beta \) – regression coefficient, and \( \epsilon \) – constant. It should be noted that here \( \epsilon \) is taken as equal to 0. Specifically, for the model developed here the multiple regression model is:

\[ \ln(X) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(Y) + \beta_3 \ln(T) + \beta_n \ln(p_B) \]  

(5.18)

Or more conveniently:

\[ X = e^{\beta_0} D^{\beta_1} Y^{\beta_2} T^{\beta_3} p_B^{\beta_n} \]  

(5.19)

The multiple regression analysis used in this study was carried out with MS Excel using its Data Analysis add-in. To acquire the requisite data for this analysis, every output value of interest, that is FCI and COGS, from the model was recorded for every possible combination of instances of independent variables, that is \( D, Y, T, \) and \( p_B \), within the desired range of values (Table 5.5). In total this constituted 1680 combinations of instances of independent variables each requiring an output value. This process was automated using VBA macros in MS Excel. All input and output values were subsequently converted to their natural logarithm equivalent and processed using linear regression, which allows determination of the geometric relationship.
In addition to the model, the coefficient of determination, $R^2$, was recorded to measure the significance of the independent variables in determining the dependent variable. $R^2$ is bound between 0 and 1. A high $R^2$ value means that a large proportion of the variance can be actually attributed to the independent variables as established and explained by the learned model. A low $R^2$ value means that the correlation is relatively weak, and suggests a poor ability of the independent variables to explain the dependent variable. Therefore, $R^2$ represents the percentage of the variance in the dependent value that can be solely attributed to the effects of the independent variables within the dataset analysed. By comparing different $R^2$ values one it can be determined where the model produces acceptable accuracy within the ranges of input values considered.

5.3. RESULTS

The results are now presented and discussed.

5.3.1. FIXED CAPITAL INVESTMENT

Figure 5.4a and Figure 5.4b both show that total fixed capital investment (FCI) increases with increasing annual output. This is expected because with increasing annual output, processing equipment with greater capacities and greater numbers are clearly required. Figure 5.5a and Figure 5.5b complement these figures by demonstrating how behaviour in FCI is influenced by behaviour in equipment costs from the upstream processing (USP) and downstream processing (DSP) streams. This is because FCI is calculated as a linear function of equipment cost via the Lang factor method. Figure 5.4a shows that there are regions where the relationship between FCI and annual demand shows gradual progression with intermittent accelerated increases. These intermittent accelerated increases are seen at annual demand ranges of 100-125kg, 150-175kg, 225-250kg, 300-325kg, 350-375kg, and 475-500kg. Analysis shows that these increases are due to requirements for additional equipment. The need for additional equipment occurs at the upper bound of each of these ranges because annual demand is sampled at every 25kg. Increases in FCI outside of these ranges are attributable to increasing equipment sizes within commercially available...
dimensions where there is no need for increased equipment numbers. Such regions are seen to be 175-225kg and 375-475kg. At 175kg, 325kg, and 500kg an additional fermenter is added to the facility whose magnitude of expense causes an accelerated increase in FCI at each of these demand levels. These increases simply indicate that it is more expensive to increase the number of fermenter units in the production process than it is to host the entire volume of multiple fermenters in a single larger fermenter when possible. Other additional USP equipment are not required in the demand range considered because of the smaller process volumes that they handle. In Figure 5.5a it can be immediately seen that the range for DSP equipment cost is larger than that for USP equipment cost, and the behaviour of DSP equipment cost against demand is generally more abrupt than with USP equipment cost. This is due to the greater number of DSP operations than USP operations combined with the significant cost of DSP equipment units. Resultantly, potentially more equipment will have to be scaled in appropriate numbers to meet demand levels. Table 5.6 details which additional equipment is required with increasing annual demand. Additions to DSP equipment numbers occur at a faster rate than with USP equipment numbers because of the greater number of DSP operations.

![Figure 5.4](image)

**Figure 5.4.** Fixed capital investment (£MM) against annual demand for production of: (a) mammalian cell-culture derived monoclonal antibodies, (b) *E.coli* derived antibody fragments. Titre = 1g/L, yield = 0.6, and *p* = 0.9.
ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY MANUFACTURE AND ANTIBODY FRAGMENTS MANUFACTURE

Figure 5.5. Fixed USP (-----) and DSP (-----) equipment costs (£MM) against annual demand for: (a) mammalian cell-culture derived monoclonal antibodies, (b) E.coli derived antibody fragments. Titre = 1g/L, yield = 0.6, and \( p_b = 0.9 \). Equipment cost figures include all equipment required for unit operations and their ancillary equipment such as holding tanks.

Figure 5.4b has a very different profile to that seen in Figure 5.4a. For annual demands below 150kg the capital expense for the bacterial derived manufacture of antibody fragments is greater than that for the mammalian derived manufacture of whole monoclonal antibodies. Figure 5.5b shows that this is due to the greater DSP equipment cost for the bacterial derived manufacture of antibody fragments. This is intuitive as the process for producing antibody fragments has a downstream process that is 5 operations longer than for the manufacture of whole antibodies. Thus the fixed capital expense will be greater at lower demands where only one equipment unit is required for each unit operation. At demands greater than 150kg the capital expense required for the manufacture of antibody fragments is consistently lower than that for whole monoclonal antibodies. This crossover is explained by the comparatively low processing volumes seen in the bacterial derived manufacture of antibody fragments and how this influences the rate at which equipment additions are made, and hence the rate of increase of FCI. Specifically, the manufacturing process for antibody fragments produces 50 batches per year which is 2.5 times more than for the mammalian cell-based process. Hence for the same annual output, the antibody fragments facility has a lower output per batch and smaller process volumes per batch when compared to the mammalian derived whole monoclonal antibody manufacturing facility. Overall for bacterial derived antibody fragments manufacture, FCI is higher at annual demands below 150kg because of the greater equipment numbers utilised when only one equipment unit is required for each unit operation in either manufacturing process; whilst FCI is lower above annual demands of 150kg because
of the significantly lower rate at which equipment numbers are increased to meet the annual demand relative to mammalian derived whole monoclonal antibody manufacture. Like Figure 5.4a, regions of gradual increase are seen with regions of intermittent accelerated increases. Figure 5.4b has a significantly more prolonged initial period of gradual increase, which is the region 0-375kg. This is because within this range equipment sizes can be increased without exceeding commercially available equipment dimensions. At 400kg an ultrafiltration unit is added to the facility, and at 425kg an additional fermenter is added.

For the mammalian derived monoclonal antibody manufacturing process, FCI values bear similarity to published figures in Farid (2007), but appear to lean towards underestimation. Farid (2007) includes a survey of recently completed biopharmaceutical manufacturing facilities where FCI figures fall within the range of $50MM-$600MM. Understandably it is difficult to achieve a complete analysis as the manufacturing procedures, technical manufacturing performance metrics, and exact facility design are all unknown for these facilities. FCI values specifically for bacteria derived antibody fragments have not been found in literature.

<table>
<thead>
<tr>
<th>Annual Demand (kg)</th>
<th>Mammalian derived whole monoclonal antibody manufacture</th>
<th>Bacteria derived antibody fragments manufacture</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>Ultrafiltration unit</td>
<td>-</td>
</tr>
<tr>
<td>175</td>
<td>Fermenter</td>
<td>-</td>
</tr>
<tr>
<td>225</td>
<td>Cation-exchange chromatography column</td>
<td>-</td>
</tr>
<tr>
<td>250</td>
<td>Ultrafiltration unit</td>
<td>Affinity chromatography column</td>
</tr>
<tr>
<td>325</td>
<td>Fermenter</td>
<td>-</td>
</tr>
<tr>
<td>375</td>
<td>Ultrafiltration unit</td>
<td>-</td>
</tr>
<tr>
<td>400</td>
<td>-</td>
<td>Ultrafiltration unit</td>
</tr>
<tr>
<td>425</td>
<td>-</td>
<td>Fermenter</td>
</tr>
<tr>
<td>500</td>
<td>Fermenter</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Equipment required for a level demand builds upon those required at previous levels. '-' indicates that no equipment additions were made at the referenced annual demand level.
5.3.2. FERMENTATION AND CHROMATOGRAPHY CAPACITY

Figure 5.6a and Figure 5.6b demonstrate how the total fermentation and chromatography capacities scales with annual demand when all other factors are constant. It is observed that for mammalian derived whole monoclonal antibody manufacture both fermentation and chromatography capacities are consistently approximately 2.5 times greater. This reflects the greater processing volumes required by this manufacturing process as it runs 2.5 times less batches per year when compared to the manufacture of bacteria derived antibody fragments. It is also seen that chromatography capacity scales at an increased rate for annual demands above 200kg because of the additional chromatography units required by the facility. For bacteria derived antibody fragments manufacture this relationship scales more linearly as there are no additions to chromatography equipment made within the range of annual demand values considered. In both cases fermentation capacity scales almost linearly with annual demand. This is expected as fermentation capacity is calculated as a linear function of annual demand which is then rounded to match commercially available denominations of equipment dimensions.

Figure 5.6. Fermentation capacity (---) and chromatography capacity (-- - ) requirements against annual demand for: (a) mammalian cell-culture derived monoclonal antibodies, (b) E.coli derived antibody fragments. Fermentation capacity is the total fermentation capacity required across all fermenters. Chromatography capacity is the volume of all resins used across all chromatography unit operations. Titre = 1g/L, yield = 0.6, and $p_B = 0.9$.

5.3.3. FIXED INVESTMENT COST PER LITRE

Figure 5.7a and Figure 5.7b both show that FCI/L decreases with increasing fermentation capacity demonstrating that there are economies of scale with FCI/L. It
is observed that FCI with the whole monoclonal antibody manufacturing process building fermentation capacity is more efficient with regard to FCI/L. This is understandable as the output of the antibody fragment facility per batch is significantly lower than that for the whole monoclonal antibody process, and requires more DSP operations. For the mammalian-derived monoclonal antibody manufacturing process model, estimated FCI/L figures for fermentation capacities above 10,000L align with published values for mammalian derived whole monoclonal antibody manufacturing facilities. For example, work by Farid (2007) suggests that benchmark FCI/L values for antibody manufacturing facilities are in the range $1765 \leq \text{FCI/L} \leq $4220, which is approximately £900 \leq \text{FCI/L} \leq £2100.

Figure 5.7. Fixed capital investment per litre of fermentation capacity against fermentation capacity for: (a) mammalian derived whole monoclonal antibodies manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = Ig/L, yield = 0.6, and $p_B = 0.9$.

5.3.4. COST OF GOODS SOLD PER GRAM

Figure 5.8a and Figure 5.8b show how COGS/g values vary with demand for both manufacturing processes. It is expected for COGS/g to decrease with increasing annual demand because relevant fixed costs are spread over a larger annual output. Thus, each gram becomes less expensive to produce. Additionally, if the annual demand is large enough then the COGS/g should converge towards a value as the composition of COGS/g becomes increasingly based on variable costs, such as the cost of raw materials, and decreasingly based on fixed costs. Some fluctuations in COGS/g will occur with the requirement for additional equipment items. This is because equipment additions can give rise to step increases in both raw materials
costs, due to step increases in the use of consumable materials, and indirect costs, due to step increases in FCI and facility size. As observed in both cases COGS/g decreases rapidly at first, then gradually increases, and then gradually decreases. With the manufacture of antibody fragments this is far more marginal. Increases occur at annual demands of 225kg in Figure 5.8a and 425kg in Figure 5.8b which both accompany the addition of a fermenter which causes a step increase in FCI and adds significantly to indirect costs. This is because indirect costs increase with increasing capital expense and also with the increasing facility size required for accommodating additional equipment. In the manufacture of whole monoclonal antibodies a cation exchange chromatography column is also added at an annual demand of 225kg which assists in explaining why the increase in COGS/g is more dramatic in Figure 5.8a than in Figure 5.8b. The addition of any new equipment can introduce a significant step-wise increase in the use of materials, consumables, and utilities. This is particularly the case with chromatography columns because of the ancillary holding tanks and the use of expensive buffers that accompany additional columns.

Cost of goods figures for specific processes with specified technical details on manufacturing are difficult to locate in literature. Werner (2004) calculates that for a mammalian whole monoclonal antibody production facility where \( D = 250\text{kg/yr}, T = 1\text{g/L}, \) and \( Y = 0.7, \) COGS/g is $260/g or approximately £130/g. Assuming that Werner (2004) uses \( p_B = 1, \) the model in this work estimates COGS/g = £150/g using these figures. Young et al. (1997) estimate that for a mammalian production process with an annual demand of 100kg/yr the COGS/g range will be $300\leq\text{COGS/g}\leq$3000, or approximately £150\leq\text{COGS/g}\leq£1500. Even using assumptions that would have been optimistic at the time of that study, that is \( T = 1\text{g/L}, Y = 0.7, \) and \( p_B = 1, \) the model used here estimates that COGS/g = £190/g. Model assumptions in those studies are likely to have different details than those found here but it has been demonstrated that the model estimates are within the range of those seen in Young et al. (1997) and Werner (2004). Comparable figures for bacteria derived production of antibody fragments have not been found in literature.
ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY MANUFACTURE AND ANTIbody FRAGMENTS MANUFACTURE

Figure 5.8. COGS/g against annual demand for: (a) mammalian derived whole monoclonal antibody manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = 1g/L, yield = 0.6, and \( p_B = 0.9 \).

From observing Figure 5.9a and Figure 5.9b it is clear that the cost of direct raw materials is by far the main driver of COGS/g, except at low levels of annual demand where indirect costs prevail. For both processes, the percentage comprising direct raw materials increases significantly from low levels of demand upwards where the rate of increase consistently decreases. This is because raw material costs are variable costs which must increase with the increased production requirements. These costs increase at a faster rate than indirect costs because the use of raw materials is more positively correlated to annual demand than FCI. Labour costs comprise the most modest percentage of COGS/g because amongst the type of costs considered it is the least correlated to annual demand. Also, the need for staffing is most correlated to the number of equipment units in the facility. The sharp increase in the direct raw materials percentage in Figure 5.9a at 225kg is primarily associated with the addition of a cation exchange chromatography column which uses expensive buffers. The accelerated increases in indirect cost percentage observed in Figure 5.9a at 125kg, 250kg, 375kg, and 500kg are primarily associated with the addition of ultrafiltration units which adds to indirect costs because of considerable increases in FCI and facility size. The increase in the indirect costs percentage seen in Figure 5.9b at 400kg is associated with the addition of an ultrafiltration unit.
Figure 5.9. The percentage that direct raw materials(——), indirect costs (-----), and labour (-----) comprise COGS/g against annual demand for: (a) mammalian derived whole monoclonal antibody manufacture, and (b) bacteria derived antibody fragments manufacture. Titre = 1g/L, yield = 0.6, and $p_B = 0.9$.

5.4. RESULTS: ECONOMIC APPROXIMATIONS

The approximations derived from multiple regression analysis are now presented. The reader is reminded that these approximations take the form:

\[
\ln(X) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(Y) + \beta_3 \ln(T) + \beta_4 \ln(p_B) \\
X = e^{\beta_0} D^{\beta_1} Y^{\beta_2} T^{\beta_3} p_B^{\beta_4}
\]

Where units are: $D$ – kg, $Y$ – percentage, $T$ – g/L, $p_B$ – percentage, $FCI$ – £, and $COGS$ – £/g. Tables 5.7 through 5.10 give the regression coefficient, coefficients of determination, t-statistics, and p-values for specified ranges of annual demand. Region specific coefficients have been investigated so as to improve accuracy of the multiple regression approximation. The specified ranges have been chosen according to the behaviour of FCI and COGS/g against annual demand. It can be observed that for each region a different set of values for regression coefficients has been derived showing that, as expected, a ubiquitous model would be less accurate than a region specific model. Coefficients of determination increase with increasing annual output due to the less abrupt changes in the output values. Generally, high coefficients of determination are observed for most regions considered. The coefficients of determination show considerably less accuracy in the region $0 \text{kg/yr} < D \leq 100 \text{kg/yr}$ when compared to regions of higher demand for which more accurate approximations are derived. This is intuitive as this region is the most sensitive to changes in the
factors considered. It is also possible that because of this, step changes in output values are seen towards the upper bound for the annual demand and the lower bounds for the remaining factors. As the level of demand increases the behaviour of FCI becomes increasingly less abrupt within the ranges considered. It can be seen that for both FCI and COGS/g, coefficients of determination are generally above 0.9 for the range $100\text{kg/yr}<D<500\text{kg/yr}$. The t-statistics are generally high for all ranges of demand suggesting that the regression coefficients observed are statistically different from a coefficient value of zero. Additionally, the p-values generally indicate that the t-statistics are significant to at least the 0.1% level. This shows that there is at most a probability of 0.001 that the relevant t-statistic could be the value observed if the true coefficient is actually zero. These observations strongly suggest that the coefficients are valid for the range of demand values considered. The only exception is $\beta_1$ for the estimation of COGS/g for antibody fragments manufacture in the range $100\text{kg/yr}<D<500\text{kg/yr}$. For this particular coefficient the p-value is only significant to the 10% level. While this itself does not invalidate the coefficient it does show that, for this range of demand values, demand has less of a statistically significant impact on estimating COGS/g for antibody fragments manufacture than the other variables considered. For FCI concerning the manufacture of mammalian cell culture derived whole monoclonal antibodies, the derived correlation is in the range $100\text{kg/yr}<D<500\text{kg/yr}$:

$$FCI = e^{11.78 D^{0.98} Y^{-0.97} T^{-0.97} p_B^{-0.99}}$$

Interestingly, as the variable exponents are all close to 1 it indicates that the FCI for whole monoclonal antibody manufacture is almost directly proportional to $D$ and almost inversely proportional to $Y$, $T$, and $p_B$. For FCI concerning the bacteria derived antibody fragments manufacturing process, the derived correlation is in the range $100\text{kg/yr}<D<500\text{kg/yr}$:

$$FCI = e^{13.91 D^{0.58} Y^{-0.60} T^{-0.63} p_B^{-0.69}}$$

166
### Table 5.7. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of FCI for whole monoclonal antibody manufacture.

<table>
<thead>
<tr>
<th>Annual Demand Region</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0kg/yr &lt; $D$ ≤ 100 kg/yr</td>
<td>coefficient: 16.225</td>
<td>0.185</td>
<td>-0.253</td>
<td>-0.244</td>
<td>-0.249</td>
<td>0.659</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 327.5</td>
<td>17.2</td>
<td>-5.9</td>
<td>-20.7</td>
<td>-3.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; $D$ ≤ 350 kg/yr</td>
<td>coefficient: 12.098</td>
<td>0.937</td>
<td>-0.881</td>
<td>-0.853</td>
<td>-0.922</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 100.2</td>
<td>42.9</td>
<td>-26.0</td>
<td>-91.1</td>
<td>-17.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>350kg/yr &lt; $D$ ≤ 500 kg/yr</td>
<td>coefficient: 10.950</td>
<td>1.095</td>
<td>-1.101</td>
<td>-1.177</td>
<td>-1.103</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 21.8</td>
<td>13.2</td>
<td>-28.3</td>
<td>-109.5</td>
<td>-18.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; $D$ ≤ 500 kg/yr</td>
<td>coefficient: 11.778</td>
<td>0.981</td>
<td>-0.964</td>
<td>-0.974</td>
<td>-0.990</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 130.9</td>
<td>63.6</td>
<td>-31.6</td>
<td>-115.8</td>
<td>-20.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 5.8. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of FCI for antibody fragment manufacture.

<table>
<thead>
<tr>
<th>Annual Demand Region</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0kg/yr &lt; $D$ ≤ 100 kg/yr</td>
<td>coefficient: 16.975</td>
<td>0.061</td>
<td>-0.082</td>
<td>-0.078</td>
<td>-0.091</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 913.6</td>
<td>15.1</td>
<td>-5.1</td>
<td>-17.8</td>
<td>-3.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; $D$ ≤ 350 kg/yr</td>
<td>coefficient: 14.801</td>
<td>0.440</td>
<td>-0.432</td>
<td>-0.489</td>
<td>-0.485</td>
<td>0.799</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 118.4</td>
<td>19.5</td>
<td>-12.3</td>
<td>-50.4</td>
<td>-8.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>350kg/yr &lt; $D$ ≤ 500 kg/yr</td>
<td>coefficient: 11.111</td>
<td>0.997</td>
<td>-0.876</td>
<td>-0.867</td>
<td>-1.032</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 19.1</td>
<td>10.4</td>
<td>-19.4</td>
<td>-69.6</td>
<td>-14.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; $D$ ≤ 500 kg/yr</td>
<td>coefficient: 13.905</td>
<td>0.578</td>
<td>-0.598</td>
<td>-0.631</td>
<td>-0.690</td>
<td>0.821</td>
</tr>
<tr>
<td></td>
<td>t-statistic: 135.3</td>
<td>32.8</td>
<td>-17.2</td>
<td>-65.6</td>
<td>-12.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p-value: &lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 5.9. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of COGS/g for whole monoclonal antibody manufacture.

<table>
<thead>
<tr>
<th>Annual Demand Region</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0kg/yr &lt; ( D \leq 100 ) kg/yr</td>
<td>7.661</td>
<td>-0.540</td>
<td>-0.456</td>
<td>-0.330</td>
<td>-0.517</td>
<td>0.932</td>
</tr>
<tr>
<td>t-statistic</td>
<td>191.6</td>
<td>-62.5</td>
<td>-13.3</td>
<td>-34.8</td>
<td>-9.6</td>
<td>-</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; ( D \leq 500 ) kg/yr</td>
<td>3.972</td>
<td>0.171</td>
<td>-0.908</td>
<td>-0.551</td>
<td>-1.114</td>
<td>0.946</td>
</tr>
<tr>
<td>t-statistic</td>
<td>130.9</td>
<td>63.6</td>
<td>-31.6</td>
<td>-115.8</td>
<td>-20.7</td>
<td>-</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.10. Region specific regression coefficients, coefficients of determination, t-statistics, and p-values for the approximation of COGS/g for antibody fragment manufacture.

<table>
<thead>
<tr>
<th>Annual Demand Region</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0kg/yr &lt; ( D \leq 100 ) kg/yr</td>
<td>8.182</td>
<td>-0.663</td>
<td>-0.382</td>
<td>-0.394</td>
<td>-0.409</td>
<td>0.934</td>
</tr>
<tr>
<td>t-statistic</td>
<td>174.5</td>
<td>-65.4</td>
<td>-9.5</td>
<td>-35.4</td>
<td>-6.5</td>
<td>-</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>100kg/yr &lt; ( D \leq 500 ) kg/yr</td>
<td>4.564</td>
<td>-0.021</td>
<td>-0.952</td>
<td>-0.998</td>
<td>-1.055</td>
<td>0.952</td>
</tr>
<tr>
<td>t-statistic</td>
<td>64.9</td>
<td>-1.7</td>
<td>-40.0</td>
<td>-151.6</td>
<td>-28.2</td>
<td>-</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>0.088</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
</tbody>
</table>

The error analyses for the multiple regression approximations for mammalian-derived whole monoclonal antibody manufacture are seen in Figure 5.10 and Figure 5.11. It can be seen that the error does not exceed 0.40 for the fixed capital expenditure correlation. The majority of the error values are equal to or below 0.20. The approximation for COGS/g generally shows a more favourable margin of error, which does not exceed 0.20 for the vast majority of input variables considered. In Figure 5.11 a band of increased error is seen for the range 100kg/yr < \( D \leq 250 \) kg/yr. This is attributed to the step changes that accompany the need for additional process equipment.
Analyses of the scope of error values seen for FCI approximations of bacteria-derived antibody fragment are demonstrated in Figure 5.12 and Figure 5.13. For the majority of input values for FCI approximation it can be seen that the magnitude of error is within 0.20. A similar behaviour in the margin of error as it varies against annual demand and yield is exhibited between Figure 5.10 and Figure 5.12, where there is a central diagonal band of relatively high error. These are attributed to the step changes that accompany the need for additional process equipment. The COGS/g approximations show an error that is mostly equal to or less than 0.20.
Conclusions

Approximations for the capital expense and COGS/g have been developed for construction of facilities for whole monoclonal antibody manufacture and antibody fragment manufacture. A mathematical model was developed specific to each manufacturing process so that the inputs of annual demand, yield, titre, and batch success probability could be used to generate output metrics that would be of interest to decision makers in the corresponding industry. It has further been shown that the model outputs can successfully be approximated using multiple regression analysis. In this study a geometric regression model was used and has been shown to produce acceptable levels of accuracy when compared to actual model outputs.
CHAPTER 6

COMMERCIAL CONSIDERATIONS

6.1. INTRODUCTION

In previous sections the theoretical development of a computational framework for the stochastic, combinatorial, and multi-objective optimisation of biopharmaceutical portfolio management and capacity acquisition strategies has been described in full detail. As a reminder the framework has been set up into two broad areas: one which uses Monte Carlo simulation to evaluate the stochastic performance of strategies, and another that optimises the strategies according to pre-defined objective criteria. Here the focus has been on acquiring a set of computational tools that together form a process that is fit for the purpose described. For the framework to be transformed into a software tool of commercial standards and expectations the focus of development must be shifted towards catering for the needs of the end user. The details that need to be considered for such software development aims are:

- An intuitive and aesthetically pleasing graphical user interface (GUI).
- Definable biopharmaceutical development environment so that the user can present an accurate computational description of the current corporate and technological situation to the software tool.
- Objective criteria that can be selected by the user.
- Fully adjustable settings for the simulation and optimisation routines such that the robustness of the solution and the timescale for the routine can be controlled by the user.
- Professionally formatted output reports and datasets that are exportable for use with popular data analysis software packages.
- Multi-threading options so that routines can be performed with multiple processors for increased efficiency.
- An adequately descriptive help file that includes troubleshooting advice and aligns with the expectations of the most stringent regulatory bodies.
The following section explores these features in further detail alongside how they are connected within the software architecture.

### 6.2. Software Architecture

![Software Architecture Diagram](image)

Figure 6.1. Schematic of software architecture.

Figure 6.1 demonstrates the principal elements and interconnectivity of the software architecture. It is recommended that in the interests of time efficiency that the model is built within a single fast execution programming language such as C++, or a programmable environment such as Matlab™. Because of the intense computational requirements it is imperative that the programming language or environment provides fast execution speeds and multi-threading capabilities for the use of multiple processors. Additionally, some form of visual output for intermittent and final results is seen as a desirable feature for user convenience. The end users are considered to be but not restricted to:

- Technical support staff and consultants of the firm issuing this software.
- Decision-making specialists in industry.
- Management consultants that have a biopharmaceutical focus.
- Venture capitalists who are sufficiently involved in the direction of a biopharmaceutical start-up company.
- Academics.
COMMERCIAL CONSIDERATIONS

- University and business school research teams.

In each case it is expected that the user will be familiar with:
- Biopharmaceutical manufacturing and economics.
- Monte Carlo sampling.
- Evolutionary computation.
- Probabilistic modelling using Bayesian networks.
- Data analysis techniques.

The benefit to such clients is clear. Those aiming to use the software for commercial insight can benefit from the data produced as well as the strategies discovered by the framework. Significant cost savings or profit generation can result from having an improved set of strategic decisions. At present, because such frameworks do not appear to be used in the industry it provides an opportunity for business creation by software developers who can also act as full service providers with employed consultants. Those involved in academic research can use the tool for economic analysis of commercial scenarios. The following sections detail the principal elements of the model architecture and how the software may be used in a client and consultant commercial relationship.

6.2.1. Graphical User Interface

The GUI design should be intuitive enough to facilitate rapid learning of the software environment as well as principal functions and settings. The GUI design features should include:
- Drop-down menus that are categorised and sub categorised for ease of use.
- Menu bars that can be docked and floated.
- Access to all settings of the modelling environment, simulation protocols, optimisation protocols, and database.
- Convenient visualisation of numerical and graphical data.
- Data input from previous sessions.
- Add-on functionalities which are easy to locate and use.

Drop-down menus should provide convenient access to the functionality of the software and assist in facilitating the speed at which users can meaningfully use the software as an advanced analytical tool. Additionally, sensible design of right-click accessible features from the mouse should be applied alongside sensible functional
design for use of the left mouse button. An added convenience is when menu bars that are important to the user can be accessed by floating them temporarily or docking them in the standard menu bar. Access to all settings would allow for the design of a computational process that is fully definable by the user. This feature allows the user only make visual the tools that they are likely to require. The ability to conveniently access and visualise data inputted to the model and data produced by the model is very important as the model is expected to produce very large datasets. It should be highlighted that the software should still be able to do this while simulation and optimisation procedures are under execution so sufficient computational and memory related resources should be dynamically allocated for this. Data input from previous sessions would allow for a session to be stopped and restarted at some future point. Finally, further features will undoubtedly be produced as a result of individualistic creation via the software design firm and may be in the form of updates, patches, and add-ons that provide extended specialist functionalities. All such software items must be straightforward to incorporate. In the case where extended functionalities are provided then these features need to be easily found and applied by the user once the installation procedure has been completed.

6.2.2. Modelling Environment

The modelling environment should be a flexible environment that can define any process pertinent to pharmaceuticals or biopharmaceuticals to include:

- Manufacturing processes.
- Cost databases for raw material, utilities, and staffing.
- Commercial characteristics of drugs.
- Dependency relationships between options made within a portfolio of drugs\(^1\).
- Cost information on third parties.
- Costs and transition probability data related to the drug development process.

A visual tool should be used for defining the manufacturing processes within the modelling environment. This procedure should use icons that represent unit operations and their relation within the manufacturing environment. A considerable range of manufacturing processes should already be provided for access within the

\(^1\) See chapter 3 for more information on dependencies.
software database. The user should be able to change any settings within the provided manufacturing process. Adjustable settings would include mass balancing, costing, and staffing information. Importantly, there should be a design feature where the user can define new unit operations and create bespoke manufacturing processes. All cost databases should be amenable for extension according to the user's databases.

6.2.3. Simulation Protocols

The simulation protocols control what is simulated and how it is simulated. This should be featured as an accessible list that shows:

- Whether an input variable is deterministic or stochastic.
- Distribution types and numerical settings of stochastic variables.
- The Monte Carlo sampling protocol.

The variable type and stochastic features should be easy to represent within a single page. The distribution types can include triangular, Gaussian, and log-Gaussian types. The Monte Carlo sampling protocol should include a range of advanced features such as stratified sampling, quasi-random Monte Carlo sampling, and Markov Chain Monte Carlo, for example. Simulation protocols should be executable independent of the optimisation protocols. This would allow for use of the software as a calculation tool for determining important characteristics of manufacturing processes and company cashflows.

6.2.4. Optimisation Protocols

User definable features of the optimisation protocols include:

- Decision variables that comprise a strategic superstructure.
- Instances of each decision variable.
- Number of Monte Carlo samples that should be used for each strategy.
- Population size.
- Objectives.
- Number of top performing strategies that will be selected for probabilistic modelling.
- Bayesian network modelling settings.
Decision variables should be restricted to predefined types which are the drugs available for selection, the corporate body that completes the activity, and the scheduling strategy. The remainder of the above list can be made adjustable by presenting these variables as a combination of lists from which options can be selected and input boxes where required numerical data can be entered. A list of predefined objectives should be included with the software. Bayesian network modelling settings will include restrictions, if any, on the number of parents that a node can have or if a particular decision variable can never be modelled as being probabilistically influenced by another decision variable.

6.2.5. DATABASE

The database will include reference items such as cost items but will also include information on results. Results databases will have to be segregated into the generation of the evolutionary computation procedure by which they are generated. Also, the user should be able to define what is recorded with minimal or no restrictions.

6.2.6. ANALYTICS

A reasonable range of data analytics should be included in the software package alongside a range of graphing options. The data should be exportable such that it can be used in other data analysis and graphing packages. One important function to consider would be the type of analytics seen in chapter 5 where the outputs of a particular manufacturing process can be modelled according to relationships derived by regression analysis. Provided that the use of regression relationships would fall within acceptable margins of error then these relationships could be used to replace the need for using the computationally expensive models. Significant savings on computational requirements will give increase the time efficiency by which any results are obtained.
6.3. PROJECT IMPLEMENTATION

It would be unlikely that the software could be sold without the need for heavy technical support and consultation from the developer. Furthermore, it would be imprudent not to take advantage of opportunities to charge consultation fees for bespoke use of the software. This section details the process by which the software is intended to be used within a commercial project planning environment.

6.3.1. PHASE I: ANALYSIS OF USER REQUIREMENTS

- Consultants visit client site in order to gain an understanding of the user requirements and goals.
- Elements of the tool that require bespoke modelling will be identified for further attention.
- Computer hardware available at the client site will be assessed for compatibility with the software to be delivered.
- A realistic timeline for project completion to include all relevant tasks will be proposed, negotiated, and agreed upon.
- A contract will be agreed upon to address issues that include payment, work to be completed, specified users, the software license, and expiry date.

6.3.2 PHASE II: SYSTEM DESIGN AND DEVELOPMENT

- Consultants visit client site to determine a full bespoke list of modelling and simulation requirements.
- Fully functional software designed for specific use by the client is developed at the developer’s site.
- Any data that is likely to change with time is validated using the relevant sources.
- All documentation is validated for accuracy.
- The optimal hardware configuration for use with the software should be developed on site for expedient processing of client settings. The hardware is most likely to be a grid with several high-end processors to take full advantage of the multi-threading functionality of the software. This presents an opportunity for the company to charge a fee for its use.
6.3.3. Phase III: Operation and Maintenance

- Installation of software and additional modules at client site with the relevant license applied.
- Specified users will be trained.
- Maintenance will be maintained via online means primarily and then through visits to the client site when necessary.

6.3.4 Project Cost Analysis

An example cost analysis is provided to give some approximate means of the considerations and costs to run such a project within a commercial environment. The assumptions apply:

- Consultants command a rate of £1000 per day.
- Contractually specified duties will be completed by the agreed timeline.
- Additional services are charged on a per use basis.
- No legal fees for the construction of the contract are applied.
- No mark up is included for services rendered.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Task</th>
<th>Resource</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>Client visit</td>
<td>1 consultant for 1 week</td>
<td>£5,000</td>
</tr>
<tr>
<td>Phase II</td>
<td>Client visit</td>
<td>1 consultant for 1 week</td>
<td>£5,000</td>
</tr>
<tr>
<td></td>
<td>Development of bespoke features</td>
<td>2 consultants for 8 weeks</td>
<td>£16,000</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>1 consultant for 2 weeks</td>
<td>£10,000</td>
</tr>
<tr>
<td></td>
<td>Documentation</td>
<td>1 consultant for 2 weeks</td>
<td>£10,000</td>
</tr>
<tr>
<td>Phase III</td>
<td>Client visit for software installation</td>
<td>1 consultant for 1 week</td>
<td>£5,000</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>1 consultant for 1 week</td>
<td>£5,000</td>
</tr>
<tr>
<td></td>
<td>Unrestricted software license</td>
<td>-</td>
<td>£10,000</td>
</tr>
<tr>
<td></td>
<td>Total Cost</td>
<td></td>
<td>£66,000</td>
</tr>
<tr>
<td></td>
<td>Total Duration</td>
<td></td>
<td>16 weeks</td>
</tr>
</tbody>
</table>

Figure 6.2. Cost and time breakdown for an example project.

A cost analysis of an example project is shown in Figure 6.2. Additional services to be charged for include:
- Additional client visits for follow-on consultation in the interpretation of results and assessment of new or modified facilities.
- Modification of licensed software modules.
- Additional software modules.
- Use of developer’s hardware for expedient results.
CHAPTER 7

VALIDATION

7.1. INTRODUCTION

In the biopharmaceutical industry any process that impacts the manufacture of final product must be validated by law according to the guidelines of the respective regulatory body where it will be licensed and used. The FDA provides full information on the validation of products and processes. Most applicable for the future uses of any software items based on this work are the FDA regulations for validation and validation protocol (Food and Drug Administration, 2007a) as well as software verification and validation (Food and Drug Administration, 2007b). Validation is defined by the FDA as establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its pre-determined specifications and quality attributes. Validation protocol is defined by the FDA as a written plan stating how validation will be conducted, including test parameters, product characteristics, production equipment, and decision points on what constitutes acceptable test results. Such validation processes should be catered for by the software developer so that ease of validation by the client is facilitated. For instance, the software developer should ensure that documentation aligns with the most stringent expectations of all regulatory bodies and that no part of the software violates any regulations drafted by any regulatory body. Such considerations by the software developer may also have the added benefit of encouraging uptake of the software and associated services. Alongside being a legal requirement the process of validating software in the context of this work can have the following benefits:

- Increased usability and reliability of the software, resulting in decreased failure rates, fewer recalls and corrective actions, less risk to patients and users, and reduced liability.
- Protection of commercial reputation.
A sensible long term approach to validation can reduce long term costs by making it easier and less costly to reliably modify software and revalidate software changes.

- Increased understanding of all associated systems and processes, both virtual and physical.
- Improved accuracy of modelling and optimisation procedures.
- Ensuring the safety of manufactured therapeutics.
- Ensuring commercially sensible decisions.
- Encouraging the development of appropriate and demonstrable quality control protocols and error handling mechanisms within the software tool.
- Galvanising the integrity of any data produced by the software.

The remainder of this section explores the principal aspects of software validation in further detail.

### 7.2. Software Verification and Validation

This section is based on regulations presented on software verification and validation by the FDA (Food and Drug Administration (2007b) where it makes specific references to software verification and software validation. Software verification is the provision of objective evidence that the design outputs of a particular phase of the software development life cycle will meet all of the specified requirements for that phase. Software verification looks for consistency, completeness, and correctness of the software and its supporting documentation, as it is being developed. These and suchlike activities will aim to provide support for the conclusion that software is indeed validated. Software testing is one of many verification activities intended to confirm that the software development output meets its input requirements. Other verification activities include static and dynamic analyses, inspections of codes and documents, and walkthroughs of the full function of the software. Software validation activities may occur both during, as well as at the end of software development to ensure that all requirements have been fulfilled. Since software is usually part of a larger hardware system, software validation will include evidence that all software requirements have been implemented correctly and completely, and aligns with hardware requirements. The conclusion that software is fully validated is highly...
dependent upon an appropriate plan, execution of software testing, inspections, and analyses performed at each stage of software development.

Software verification and validation in the context set out by this work can be difficult because model assumptions must be deemed acceptable for commercial use. In the worst case complex modelling of variables in the software may be required where more straightforward techniques were originally used. The FDA says that software validation is a matter of developing a level of confidence that the software meets all requirements and user expectations for its functions and features. The level of confidence, and hence the level of software validation, verification, and testing efforts required will also vary depending upon the safety risks imposed by the intended use of the software. In this case it must be assured that strategic suggestions discovered by the model do not lead a biopharmaceutical developer to configuring a strategy that is hazardous to its workers, third party associates, and ultimately the patients for which the biopharmaceutical therapeutics will be made.

The FDA recommends that the period over which a software product is developed, the software life cycle, include the following activities:

- **Quality Planning** which identifies necessary tasks, procedures for anomaly reporting and resolution, necessary resources, and management review requirements, including formal design reviews.
- **System Requirements Definition** which includes a rigorous definition of its intended use.
- **Detailed Software Requirements Specification** which details the identification, analysis, and documentation of information about the software.
- **Software Design Specification** where the software requirements specification is translated into a comprehensive physical representation of the software to be implemented.
- **Construction** where the detailed design specification is implemented as source code.
- **Testing** which entails running the software under known conditions with defined inputs and documented outcomes that can be compared to their predefined expectations.
- **Installation** where the software is tested for integrity during the installation process.
- **Operation and Support** which will also include testing at the site of the user.
- **Maintenance** includes corrective, perfective, and adaptive maintenance.
- **Retirement** where issue of the software is ceased.

Additionally, software security should be considered so that the integrity of the software is immune from the impact of software viruses.

The guidelines set out by the FDA are not exhaustive and require for additional awareness to be applied so that software products operate as intended and do not generate non-obvious and significantly adverse outputs.

### 7.3. Regulation of Decision Support Systems

The FDA guidelines appear only to require strict adherence for software products that directly impact the operation of manufacturing systems and medical devices. The work presented in this thesis might be categorised as a decision support system and does not appear to be of current regulatory concern to the FDA. The reader is reminded that the scope of use of the tool is in the generation of plans for the selection of biopharmaceutical drug candidates, scheduling of critical activities in the development of drugs, and the scheduling of third party involvement. Hence, the translation of this work into a software tool for commercial use, with original intentions intact, would not affect the way in which a manufacturing process is operated.

Despite this, the decision support software will be expected to undergo a quality of software development that has an equivalent rigour to that demanded by the FDA for regulated software products. This is certainly understandable as corporate decisions of the nature covered within the scope of this work can call for the allocation of capital that amounts to the order of billions of US dollars. Thus, the reference data and the quality of routines built into the software are far from trivial. Applying the software development cycle and related regulations underscored by the FDA will facilitate the development of a software product which clients can trust. Such a
software development cycle will also provide a solid operational foundation and commercial reputation upon which the developer can build further.
CONCLUSIONS AND FUTURE DIRECTIONS

8.1. INTRODUCTION

Developing a cogent portfolio development strategy for biopharmaceutical therapeutics far in advance of execution forces the biopharmaceutical developer to consider:

- The superstructure of decision variables that defines the scope of a strategy and its inherent decisions.
- Strategic options that exist for the development, manufacture, and marketing of therapeutics with the added consideration of strategic scheduling and third party involvement policies.
- Strategic objectives that indicate the theoretical success of strategic approaches when conducting calculative investigations.
- Physical, financial, and operational requirements of developing and manufacturing a specific therapeutic.
- Stochastic variables inherent in the development process, manufacture, and revenue expectations of specific and individual therapeutics.
- The dependency relationships that exist between the activities for development, manufacture, and marketing for a group of therapeutics.
- Control of the cost, cashflow, profitability, and probability of success for an entire biopharmaceutical drug development portfolio.
- Optimisation of the strategy superstructure given specific objectives and a model of the drug development and marketing environment.

This thesis has addressed all of these features of strategy development and their associated issues in detail. It has been shown that whilst the mathematical definitions of decision variables can be straightforward the result can be that decision spaces that are far too large to be handled manually, or even by complete enumeration computationally. Strategic objectives can be numerous but understandably in literature NPV is the metric most widely used to assess the profitability of a project or
strategy. As previously underscored, an extension to this approach will include additional objectives that cater more closely to the decision-maker's objectives. Significant computational effort may be expended in assessing the probability distributions of objective criteria for a particular strategy especially when models are accurately detailed with numerous stochastic variables. Numerous strategic objectives will create a high dimensional objective space which cannot be easily visualised and where several nondominated strategic options can result. To avoid this situation it has been shown that a practical solution can be to take advantage of MCDM to combine strategic objectives into a single metric. Another demonstrated approach is via Pareto optimality which considers multiple objectives without combining them and identifies strategies that are nondominated. Whatever the set of strategic objectives used it is observed that optimal strategies still require the allocation of significant capital and operating costs. Across the strategies demonstrated in previous chapters it is anticipated that for a portfolio of therapeutics these costs can amount to the order of billions of US dollars. Finally, optimising the strategic options to generate optimal objective criteria values is a non-trivial process that in the case of non-linear criteria and a high dimensional, multi-modal objective space may require an intelligent search procedure, as demonstrated here.

8.2. CONTRIBUTIONS OF THIS THESIS

The reader is reminded that the principal aim of this work is the creation of a computational, intelligent, stochastic, combinatorial, and multi-objective decision-support framework that can optimise a strategic map for the year-by-year planning of critical development, manufacturing, and marketing activities for an entire portfolio of biopharmaceutical therapeutics, where key related findings in the development of this framework will also be distilled. The development of such a framework has been described alongside illustrations of its intended use and the key findings of related investigations. The salient contributions of research chapters 2 through 7 will now be presented.

Chapter 2 takes the approach of investigating an extended set of objective criteria to those commonly used. The intention of this chapter was to investigate if this provided additional information that could be used to formulate strategies for the manufacture of monoclonal antibody based therapeutics. Several financial and operational criteria were considered to provide a broad and detailed analysis from financial and operational perspectives. A principal outcome of this chapter was the development of a decision-support framework to assist in the selection strategies for the acquisition of biopharmaceutical manufacturing capacity. The framework provided a structured and transparent method of analysing such scenarios through the utilisation of MCDM and Monte Carlo simulation. A hypothetical case study was formulated to demonstrate the usefulness and limitations of the framework.

In analysing the case study the weighted sum method proved to be highly suitable for data handling and for the analysis of results. The Monte Carlo simulation was valuable in highlighting the probability distributions and variance of base case values. Use of the model has highlighted that the employment of a single criterion in making strategic manufacturing decisions of this nature may not allow the decision maker to be aware of other important financial and operational criteria. However, use of multiple criteria analysed under uncertainty provided an approach for identifying and confirming the optimal strategy under consideration. The approach highlights the complexity that can be involved in making decisions similar to the one analysed. Ultimately, a thorough and accurate analysis of financial and operational data is essential to make confident distinctions between feasible and attractive strategic options.

Chapter 3 takes a significant step forward from chapter 2. Chapter 2 featured a decision-support framework based on MCDM illustrated through a scenario analysis case study. Intuitively a scenario analysis based approach does not usually consider all possible options but can be useful for gaining detailed information about the functionality and scope of use for a decision-support technique. The overarching result is the development of a computational, intelligent, stochastic, multi-objective, and combinatorial optimisation framework to assist decision-making in strategies for portfolio management, clinical development, and the acquisition manufacturing capacity for biopharmaceutical therapeutics has been presented. Due to the complexity of the intended problem, techniques from artificial intelligence, in particular evolutionary computation and machine learning, were employed for an efficient search of the decision space and an effective traversing of the objective space. Multiple objectives were included and assessed through an extensive evaluative framework that was considerate of principles featured in chapter 2. The stochastic properties of candidate strategies were evaluated by Monte Carlo simulation. A hypothetical case study was formulated to demonstrate the capabilities of the framework. This case study showed that the decision-space presented consisted of $3 \times 10^{19}$ possible solutions and would take in the order of $1 \times 10^{14}$ to evaluate the objective space in its entirety. This immediately highlighted the complexity of decisions that a drug developer may be confronted with and that optimisation approaches offer invaluable advantages in discovering optimal strategies over those that are based on scenario analysis. Also observed is that in the presence of dependencies biopharmaceutical product development strategies in the real world may be best analysed when considering entire sets of decisions holistically rather than as a collection of individual decisions. It was seen that, as might be expected, that the two objective criteria applied, mean positive NPV and $p(\text{NPV}>0)$, are conflicting measures in the case study which created a nondominated front in the objective space. At all points along this nondominated frontier a gain in one criterion is coupled with a compromise in the other criterion. Additional observations included that it has been
demonstrated that an effective strategy for portfolio development can result in a p(NPV>0) value that is significantly greater than the probability of successful development for any singular drug in the portfolio. Hence, by considering a portfolio of multiple drugs it is possible to control its risk to some extent through careful strategic formulation, whilst this is not possible with a singular drug. Another is that while the use of NPV as an objective criterion is valuable many strategic approaches can generate similar NPV values. These strategies will often differ in the value they offer in other criteria giving additional opportunities for the comparison of strategic approaches. The learning of Bayesian networks from superior solutions presented here has been used for the deep datamining of the conditional and probabilistic nature of interactions between decision variables. This element of the optimisation framework has been shown to be effective and efficient in consistently improving the quality of strategies available in each population in regions that are not at the extreme of either objective. Their effectiveness in these regions is presumed to be due to their ability to iteratively learn and exploit the structure of the problem as noted by contributions in artificial intelligence literature. At the extreme of either objective the performance above random search needs to be further investigated to establish whether the negligible performance improvement is due to limitations inherent to the problem or to the algorithm. A noteworthy insight from using the framework is that use of machine learning has potential for future development in solving portfolio development and manufacturing capacity planning problems simultaneously.

8.2.3 STRATEGIC BIOPHARMACEUTICAL PORTFOLIO DEVELOPMENT: AN ANALYSIS OF CONSTRAINT INDUCED IMPLICATIONS ON STRATEGIES

Chapter 4 extended the use of the framework presented in chapter 3 by using it to analyse the impact of the size of the drug development portfolio and cashflow constraints on the discovery nondominated strategies. Analyses included a five drug portfolio and a three drug portfolio, as well as the application of a range of cash flow constraints. Again, mean positive NPV and p(NPV>0) were selected as the objective criteria. It was seen that the introduction of cash flow constraints can lead to a reduction in the expected rewards or probability of success of a strategy. The reduction in portfolio size from five drugs to three drugs produces the same mean positive NPV range but results in a significant reduction in p(NPV>0). Finally, the
CONCLUSIONS AND FUTURE DIRECTIONS

timing and third party strategies for each size of portfolio are very different showing that the size of the portfolio can have a specific impact on how the entire portfolio should be developed and commercialised regardless of how much capital is available. This signals to the decision maker that it may not be appropriate to apply the same development strategies to portfolios of different sizes.

8.2.4 ECONOMIC CORRELATIONS FOR WHOLE MONOCLONAL ANTIBODY MANUFACTURE AND ANTIBODY FRAGMENTS MANUFACTURE

Chapter 5 extends the use of the manufacturing models seen in chapters 2, 3, and 4 by investigating whether reliable approximations to the computationally intense manufacturing models can be derived. The principal benefits of this include building computationally efficient modelling procedures and straightforward formulas that accurately estimate the economics of manufacture for biopharmaceutical developers requiring preliminary figures. In addition to the mathematical model for the mammalian manufacture of whole monoclonal antibodies, a mathematical model for the bacterial manufacture of antibody fragments was also developed. Multiple geometric regression was the chosen procedure for learning approximations. Approximations for fixed capital investment and COGS/g were developed for the construction of facilities for mammalian derived whole monoclonal antibody manufacture and bacteria derived antibody fragment manufacture. The inputs to each model were annual demand, yield, titre, and batch success probability. These could be used to generate output metrics that would be of interest to decision makers in the corresponding industry. It has further been shown that the model outputs can successfully be approximated using regression analysis techniques. In this study approximations were shown to produce acceptable levels of accuracy when compared to actual model outputs with results falling mostly within a 30% margin of error.

8.2.5. COMMERCIAL CONSIDERATIONS FOR THE DEVELOPMENT OF A STOCHASTIC COMBINATORIAL MULTI-OBJECTIVE OPTIMISATION FRAMEWORK FOR BIOPHARMACEUTICAL PORTFOLIO MANAGEMENT STRATEGIES

Chapter 6 demonstrated that the frameworks presented in this thesis can be translated into a commercial software tool as part of a viable consultancy business. The
software architecture was detailed alongside examples of end users that may have a reasonable propensity for using it. An example of how a business could be constructed for use of the software was described alongside cost breakdown and project duration details.

8.2.6. VALIDATION

Software validation protocols were discussed in chapter 7. It was seen that regulations concerning software validation were not applicable for the type of software that would be developed as a direct translation of this work into a commercial software tool. Despite this it is certainly within the interests of the service provider of such software to follow equivalent validation protocols to software products regulated by stringent authorities such as the FDA. Engaging in activities like these is likely to galvanise client confidence and improve the quality of outcomes following decisions.

8.3 RECOMMENDATIONS FOR FUTURE WORK

Within the context of the strategic development and manufacture of a portfolio of biopharmaceutical drugs, the recommendations for future work presented here will aim to extend the use of:

- Probabilistic model-building genetic algorithms.
- Computing resources.
- Problem features.

This section addresses immediate opportunities for future investigations that possess the capacity to greatly enhance the breadth and calibre of the methods and insights provided by this thesis. With principal regard to commercial considerations, these improvements can lead to improvements in the:

- Range of addressed problems.
- Efficiency by which solutions to problems are acquired.
- Functionality and performance of the commercial software.
- Scope and quality of the business proposition.
- Reputation of the service provider as held by regulatory bodies and clients.
These improvements may also require lengthier development and validation periods, and may necessitate expensive hardware specifications for enhanced performance.

8.3.1. Extending the Probabilistic Model-Building Genetic Algorithm

Improvements in the features provided by the algorithms used in this work can enhance their capabilities and efficiency concerning the type of problems approached in this work. Such improvements may also enhance the quality of results obtained in more difficult problems.

In the EDA used in chapters 3 and 4 it was taken that mutation operators seen in conventional genetic algorithms were not required because the random sampling of learned Bayesian networks based on superior performing strategies was taken to provide an equivalent function. Although it has been shown that the EDA used here works on the example cases seen, opportunities for increasing the rate at which nondominated solutions are discovered have been considered. One approach is to induce a random point mutation on a given strategy after its evaluation and then re-evaluate the mutated strategy. If the mutant strategy yields an improvement in at least one criterion and no degradation in any other criterion then it will replace its original in the population. This approach assumes that is possible for an improvement to be yielded from a mutation in a single decision variable. Mutations may also be applied to in a similar manner to multiple decision variables. Sastry et al. (2004) take an alternative approach. As described earlier building blocks can be considered as a set of instances of decision variables that work together to support the performance of candidate solutions in the objective space. Their approach is to explicitly identify building blocks within a population using marginal product models. Marginal product models are partitions of the decision making superstructure that group together highly correlated decision variables. An enumerative greedy search procedure is run on each marginal product model with the original configuration being the top performing candidate in a population. Once this process is complete, the most beneficial instances of each marginal product model are used for the induction of point mutations. The first approach is obviously simpler than that taken by Sastry et al. (2004). Their approach assumes that mutations are more effective when considered
as the mutation of a building block rather than of an individual decision variable. It should be noted that it is possible for a single decision variable to constitute an individual though simplistic building block. The advantage of this approach over blind random mutations is that it further decomposes and exploits the structure of the problem to enhance the efficiency of the optimisation procedure.

The EDA used earlier in this work dynamically clustered the objective space into three partitions from which separate Bayesian network models could be learned. A method of automating the number of clusters that are identified has not been found in literature. Increasing the number of clusters will increase the specificity by which superior performing solutions are modelled. In so doing, the nondominated front in the objective space can be maintained more effectively and more confidence can be held in the quality of the nondominated front produced at the final generation. Investigating the impact of the number of clusters on the quality of the nondominated front at the final generation will require a supervised approach where increased partitioning is used until no improvement in the quality of the nondominated front can be observed.

The size of the population is important in allowing for the opportunity for superior strategies to be discovered. In larger populations it can be considered that a greater probability exists for the discovery of superior strategies. But for sufficiently large population sizes this opportunity does not inevitably increase with every larger population. Thus choosing a population size that is too large can expend time and computational resources unnecessarily. A procedure for automating the population size has been reported by Pelikan and Lin (2004). Here, two population sizes are maintained with one being twice the size of the other. Because increased opportunities for discovering superior solutions are provided by the larger population they propose that it is expected for the average performance of the larger population to be greater than that of the smaller population. If this holds true when executing the procedure then in the next generation the larger population becomes the smaller population and the new larger population is again twice the size of this smaller population. This continues until the average performance of each populations become equivalent. From here onwards the smaller population is maintained. One variant of this technique proposed here is to compare the average performance of only the set of
top performing strategies identified for selection, as this specifically targets the quality of superior strategies.

Improvements to the way Bayesian networks are constructed can also be investigated. In chapter 3 it was described that decision variables were time ordered so that earlier decisions could not be conditionally and probabilistically dependent on decisions to be executed later in the development schedule. It is recognised that it may be possible for a strategic decision that must be executed earlier to be made in preparation for a sufficiently important strategic decision that must be exercised later. The impact of relaxing this original topological constraint can be investigated. Another improvement that can be made to the procedure for constructing Bayesian networks is the added learning of local structure as reported by Pelikan (2005). Here highly correlated decision variables are algorithmically discovered and grouped together then treated as singular decision variables, which can improve the time in which Bayesian networks are learned as the nodes in the network are effectively reduced. For further information on the construction and evaluation of Bayesian networks with local structure see Chickering et al. (1997), Friedman and Goldszmidt (1999), and Pelikan and Goldberg (2001).

Performance evaluation of each strategic option can be computationally expensive and is the rate limiting factor in the running of the EDA used here. By applying acceptable performance estimation techniques there can be significant time savings in the execution of the algorithm. The first recommendation will require the estimation of the manufacturing model outputs, as seen in chapter 5. Such an approach will involve regression based modelling of the relationships between inputs and outputs of the mathematical model. The first step would be to decide the range of values upon which the regression technique will be performed and the boundary of error that all estimates must fall within. The inputs would have to be split into multi-dimensional regions for region specific regression modelling until all model coefficients yielded regression models that yield acceptable error margins. Depending on the complexity of the model this can be time consuming. It is also possible that user defined error margins will never be met in all instances. The end result would be a database of region specific coefficients that can be used within the same form of regression model which can be used as long as the original mathematical model is not updated. When
CONCLUSIONS AND FUTURE DIRECTIONS

referenced, the calculation procedure would need to search for the correct coefficients and then use them in the estimation formula. This is more efficient than the vast number of calculations required for deriving a singular performance output in the full mathematical model. This search and then calculate approach must be repeated for each Monte Carlo iteration. An alternative approach can be seen in Pelikan and Sastry (2004) and Sastry et al. (2004). Both use Bayesian networks as a tool for estimating the key performance metrics of a strategy. As performance evaluations are conducted they are added to a permanent database. When Bayesian network models are built from each population a model of how a performance metric varies according to the conditional relationships described by the Bayesian network is also built using this permanent database as a reference for the required performance data. These models are analogous to the conditional probability tables associated with Bayesian networks except that a performance metric is used instead. These models also use a linear estimation scheme. It can be seen that with this approach it may be possible to estimate the required data about an entire probability distribution of the performance of a strategy without the need for Monte Carlo sampling and is likely to be one of the most promising techniques for this purpose. Non-linear estimation schemes may certainly be applicable but none applied to this specific purpose have been seen in literature. Classification methods in machine learning literature such as neural networks may provide a suitable non-linear estimation technique for this purpose.

Alternatives to the EDA approach can be explored as viable core applications to the large-scale multi-objective combinatorial optimisation problems seen here and there are a vast number of methods to be considered. Such examples include GRASP (Pitsoulis and Resende, 2002), local search methods (Aarts and Lenstra, 2003) such as reactive search (Battiti, 1996) and tabu search (Glover and Laguna, 1997), annealing methods such as simulated annealing (Kirkpatrick et al. 1983) or quantum annealing (Das and Chakrabarti, 2005), and swarm intelligence methods such as ant colony optimisation (Dorigo et al., 1992) and particle swarm optimisation (Eberhart and Kennedy, 1995). It is important to note that there is no single optimal optimisation method and that different methods may be better suited some types of problem than to others. For this reason it may be most useful to view an optimisation method in terms of the features it possesses and comparing those to features that would be necessary and desirable for approaching a particular problem. Of course, it is always
worthwhile exploring whether the unique strengths of some of these methods may be combined to create enhanced combinatorial optimisation algorithms. For example, local search methods or annealing methods could be coupled with the EDA as a fine-tuning secondary search process to find non-dominated solutions within clusters of the objective space. GRASP operates on a similar principle. Reactive search methods also have the advantage of automatically fine-tuning algorithm parameters during the optimisation process, which could be a worthwhile feature to incorporate into the EDA. Another example is methods that utilise memory of the search process so far. Tabu search memorises which candidate solutions have already been visited so that they cannot be revisited and can include rules that restrict the formulation of candidate solutions with undesirable features. Swarm intelligence methods retain knowledge of current optimal solutions and use that information to influence the formulation of candidate solutions. Whilst the EDA uses probabilistic models to model superior candidate solutions, it does not retain a specific memory of current optimal solutions so as to have a direct impact on the formulation of other solutions, so incorporating memory retention features could prove advantageous. Also from the reverse perspective, many other relevant methods do not have the facility for acquiring detailed knowledge of the broad structure of the problem or specific advantageous building blocks of decisions as featured by the EDA approach. Overall, it is important to remember that there is typically a trade-off to be made when making enhancements of these natures, which may be in the form of programming complexity, computational requirements, or execution time. More importantly because these methods incorporate some set of heuristics it is useful to know which features are significantly advantageous to the problem. This may itself necessitate a process of trial-and-error.

8.3.2. Computing Resources

Computing resources offer significant opportunity for increasing the time efficiency by which results are obtained. The recommendations offered here cover the use of fast programming languages and parallel computing.

It is fully recommended that the model be exported in its entirety to a single, fast execution programming language, such as C++. Parts of the model presented in this
thesis have been implemented in MS Excel and VBA which is convenient for the visualisation and manipulation of data but is widely known to be much slower than C++.

Multi-threaded computing offers one of the greatest opportunities for increases in time efficiency depending on the number of computer processors that are available. Here each processor is assigned a task that can be completed in parallel with other tasks. An example of such a task is the performance evaluation of a population of strategies. If enough processors exist then the performance evaluation of each strategy can be individually assigned to a processor. In the model formulated previously the population size was 1000 strategies so if 1000 processors can be utilised then an improvement in time efficiency by a factor close to 1000 should be observed.

8.3.3. PROBLEM FEATURES

The types of problems that can be considered for further exploration will now be addressed.

In chapter 4 two portfolio sizes were considered: a three drug portfolio and a five drug portfolio. Wider ranges of portfolio sizes and cash flow constraints can be investigated to study further their impact on the types of strategies produced by the framework.

Flexible portfolio sizing was not included as a possible feature of the decision making process because the impact of portfolio size was of specific concern. However, this feature is relevant to the types of problems investigated. For a given decision making scenario in the model future investigations should include the possibility for the decision maker to find the optimal size of drug development portfolio within the boundaries of resource constraints. In addition to the constraints seen, time constraints should be added so that solutions found by the optimisation algorithm fall within a specified time of completion for development and marketing.
Additional drug types and their respective manufacturing routes should be considered. For example, antibody fragment therapeutics should be considered in the list of candidates. Also, disposable manufacturing routes could also be featured.

Variable initial scenarios of drugs in the model can give novel insight into the best strategic options when therapeutics are already undergoing the clinical trial process. Also, a model that determines the impact of competitor activity on the size of markets can be an additional feature.

Adding market competition as an additional consideration may influence portfolio construction towards markets that are less populated and increase pressure to complete drug development expediently.

Assumptions about manufacturing capabilities can be investigated. Based on historical developments and specific knowledge about the limitations of manufacturing technologies, the expected rate of improvements in titres and yields may be included as functions rather than point values. This would effectively provide some incentive to developing drugs later rather than sooner because of the economic benefits of more favourable titres and yield.

The objectives seen in chapters 3 and 4 can be extended. For example for purposes of investigating the scope of the objective space \( p(NPV>0) \) was used as an objective criterion. This could be modified to \( p(NPV>x) \) where the decision-maker is seeking a minimum acceptable level of NPV. This can also be further extended to \( p(NPV>x>y) \) where the decision-maker is seeking minimum acceptable levels of both profit and probability of success. For reasons of simplicity and transparency of results most of the objectives featured in chapter 2 were not used in the optimisation method featured in chapters 3 and 4. Further work should seek for a way to incorporate these.

Finally, the correlations developed in chapter 5 can be extended to include prospective titre and yield values in industry.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


208
REFERENCES


Myers J. *Economic considerations in the development of downstream steps for large scale commercial biopharmaceutical processes*. IBC: Production and Economics of Biopharmaceuticals; 2000 Nov 13-15; La Jolla, CA.


Pharmaceutical Research and Manufacturers of America (PhRMA). (1998). *R&D – The key to innovation*. Pharmaceutical Research and Manufacturers of America (PhRMA), Washington, DC.


REFERENCES


REFERENCES


## Appendix A: Supporting Data for Chapter 4

### Table A.1. Probabilities of drug selection for a five drug portfolio.

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Drug 1</th>
<th>F(0.82,M,M)</th>
<th>F(0.59,M,M)</th>
<th>G(0.61,L,M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 2</td>
<td>C(0.32,M,M)</td>
<td>C(0.35,M,L)</td>
<td>E(0.37,H,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>B(0.28,H,L)</td>
<td>A(0.36,H,L)</td>
<td>F(0.45,M,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>A(0.25,H,L)</td>
<td>J(0.34,H,H)</td>
<td>B(0.34,H,L)</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>J(0.36,H,H)</td>
<td>G(0.35,L,M)</td>
<td>D(0.28,L,M)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 2</th>
<th>Drug 1</th>
<th>F(0.44,M,M)</th>
<th>F(1.00,M,M)</th>
<th>H(0.47,M,H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 2</td>
<td>D(0.29,L,M)</td>
<td>C(0.31,M,L)</td>
<td>D(0.32,L,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>B(0.47,H,L)</td>
<td>E(0.29,H,M)</td>
<td>F(0.48,M,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>A(0.22,H,L)</td>
<td>D(0.30,L,M)</td>
<td>B(0.46,H,L)</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>E(0.24,H,M)</td>
<td>H(0.34,M,H)</td>
<td>G(0.34,L,M)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 3</th>
<th>Drug 1</th>
<th>F(0.86,M,M)</th>
<th>F(1.00,M,M)</th>
<th>D(0.63,L,M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 2</td>
<td>A(0.25,H,L)</td>
<td>G(0.28,M,L)</td>
<td>E(0.48,H,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>J(0.74,H,H)</td>
<td>D(0.24,H,L)</td>
<td>C(0.55,L,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 4</td>
<td>B(0.25,H,L)</td>
<td>J(0.43,H,H)</td>
<td>H(0.40,M,H)</td>
<td></td>
</tr>
<tr>
<td>Drug 5</td>
<td>H(0.56,M,H)</td>
<td>H(0.50,M,H)</td>
<td>F(0.48,M,M)</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The letter preceding the bracket indicates the most probable drug. The first term within the bracket is the probability by which the selection of the corresponding drug is observed within its cluster. The second inside the bracket is the demand level of the drug where: L – low demand, M – medium demand, and H – high demand. The third term is the compound annual growth rate of the drug where: L – low demand, M – medium demand, and H – high demand.
Table A.2. Probabilities of drug selection for a three drug portfolio.

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Drug 1</th>
<th>F(0.70,M,M)</th>
<th>F(0.99,M,M)</th>
<th>F(0.71,M,M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug 2</td>
<td>B(0.28,H,L)</td>
<td>C(0.19,M,L)</td>
<td>E(0.19,H,M)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>I(0.34,L,H)</td>
<td>A(0.19,H,L)</td>
<td>D(0.23,L,M)</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Drug 1</td>
<td>F(0.65,M,M)</td>
<td>F(1.00,M,M)</td>
<td>F(0.30,M,M)</td>
</tr>
<tr>
<td>Drug 2</td>
<td>B(0.30,H,L)</td>
<td>H(0.26,M,L)</td>
<td>A(0.30,H,L)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>A(0.53,H,L)</td>
<td>J(0.34,H,L)</td>
<td>J(0.32,H,H)</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Drug 1</td>
<td>C(0.63,M,L)</td>
<td>C(0.28,M,L)</td>
<td>A(0.65,H,L)</td>
</tr>
<tr>
<td>Drug 2</td>
<td>E(0.32,H,M)</td>
<td>G(0.25,L,M)</td>
<td>I(0.26,L,H)</td>
<td></td>
</tr>
<tr>
<td>Drug 3</td>
<td>A(0.86,H,H)</td>
<td>A(0.24,H,L)</td>
<td>C(0.42,M,L)</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The letter preceding the bracket indicates the most probable drug. The first term within the bracket is the probability by which the selection of the corresponding drug is observed within its cluster. The second inside the bracket is the demand level of the drug where: L – low demand, M – medium demand, and H – high demand. The third term is the compound annual growth rate of the drug where: L – low demand, M – medium demand, and H – high demand.
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

B.1 INOCULUM

The inoculum step is the first stage in upstream processing to cultivate product producing cells to the required cell concentration so that the product can be produced in its required quantity. This cell culture step is achieved in flasks. The equipment sizing equations for this step are as follows:

\[
V_{\text{inc}} = \frac{D}{1000TYn_Bp_B} \quad \text{(B.1)}
\]

\[
n_{\text{flask}} = \left\lceil \frac{V_{\text{inc}}}{V_{\text{flask}}} \right\rceil \quad \text{(B.2)}
\]

where the subscript \(\text{inc}\) specifies reference to the inoculum step, \(V_{\text{inc}}\) – processing volume (L), \(D\) – annual output of the facility (kg), \(T\) – product titre in the fermentation broth (kg/L), \(Y\) – product yield achieved after downstream processing, \(n_B\) – number of batch cycles that take place in one year, \(p_B\) - probability that a fermentation batch is run successfully, \(n_{\text{flask}}\) – number of flasks required, \(V_{\text{flask}}\) – flask volume taken as 0.5L,

The function seen above, \(\lceil x \rceil\), is the ceiling function that is the nearest integer greater than \(x\). \(n_B\) is taken as 20 batches/year for the whole monoclonal antibody manufacture, and 50 batches/year for antibody fragment manufacture.

Cost relationships:

\[
E_{\text{EQP,inc}} = E_{\text{flask}}n_{\text{flask}} \quad \text{(B.3)}
\]

\[
E_{\text{DMM,inc}} = E_{\text{media/L}}n_{\text{flask}}V_{\text{flask}} + V_{\text{inc}}n_{\text{inc}}(3E_{\text{CIP/L}} + 3E_{\text{WFI/L}}) \quad \text{(B.4)}
\]

\[
E_{\text{LIR,inc}} = (7.5E_{\text{steam/L}} + 2.5E_{\text{cv/L}})V_{\text{inc}}n_{\text{inc}} \quad \text{(B.5)}
\]

\[
E_{\text{LRB,inc}} = E_{\text{staff/hr}}n_{\text{staff,inc}}t_{\text{inc}} \quad \text{(B.6)}
\]
\[ n_{\text{staff,inc}} = \begin{cases} 2 & n_{\text{flask}} \leq 24 \\ 3 & \text{otherwise} \end{cases} \quad (B.7) \]

\[ E_{\text{OPR,inc}} = E_{\text{DRM,inc}} + E_{\text{DU,inc}} + E_{\text{LBR,inc}} \quad (B.8) \]

where \( E_{\text{inc}} \) – total equipment cost, \( E_{\text{DRM,inc}} \) – direct raw materials cost, \( E_{\text{DU,inc}} \) – direct utilities cost, \( E_{\text{CIP/L}} \) – clean-in-place buffer cost per litre, \( E_{\text{WFI/L}} \) – water for injection cost per litre, \( E_{\text{steam/L}} \) – steam cost per litre, \( E_{\text{cw/L}} \) – cooling water cost per litre, \( E_{\text{LBR,inc}} \) – total labour cost, \( E_{\text{staff/hr}} \) – cost of technical staff assumed to be, \( n_{\text{staff}} \) – number of technical staff assigned, \( n_{\text{inc}} \) – number of inoculum operations required, and \( t_{\text{inc}} \) – total processing time is taken as 346 hours for mammalian processes and 30 hours for bacterial processes to include operation and necessary cleaning cycles. The constants in equations (B.4 and (B.5 represent the volume of component required relative to the volume of inoculum used for CIP and SIP operations. The value of expense items that are constant are listed in Table B.1.

Table B.1. Selected materials and utilities costs.

<table>
<thead>
<tr>
<th>Item</th>
<th>Expense</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_{\text{media/L}} )</td>
<td>£10/L(^a)</td>
</tr>
<tr>
<td>( E_{\text{CIP/L}} )</td>
<td>£1.50/L(^a)</td>
</tr>
<tr>
<td>( E_{\text{WFI/L}} )</td>
<td>£0.50/L(^b)</td>
</tr>
<tr>
<td>( E_{\text{steam/L}} )</td>
<td>£0.025/L(^b)</td>
</tr>
<tr>
<td>( E_{\text{cw/L}} )</td>
<td>£0.001/L(^b)</td>
</tr>
<tr>
<td>( E_{\text{staff/hr}} )</td>
<td>£12.50/hr(^b)</td>
</tr>
</tbody>
</table>

Sources: \( a \) – personal communication with researchers at the ACBE, UCL (2004); \( b \) – Farid (2001)

**B.2 SEED FERMENTATION**

The seed fermentation steps ramp up cell concentration in small fermenters before being transferred to the commercial scale fermenter where the product is produced in crude form. This procedure is modeled as two steps in series thus at least two seed fermenters are required for this process. The first seed fermenter processes one-hundredth of the volume processed by the commercial scale fermenter, whilst the second hosts one tenth of this volume. Thus, the volume is scaled up by a factor of 10 for each operation when progressing from seed fermentation to commercial scale fermentation.
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

Equipment sizing equations:

\[
V_{sf,1,b} = \frac{D}{100TYn_B P_B} \quad (B.9)
\]

\[
V_{sf,2,b} = \frac{D}{10TYn_B P_B} \quad (B.10)
\]

\[
V_{sf} = \left[ \frac{D}{TYn_B P_B V_{sf,d} q} \right] V_{sf,d} \quad (B.11)
\]

where the subscript \( sf, sf1, \) and \( sf2, \) respectively denote reference to any seed fermentation step, the first seed fermentation step, and the second seed fermentation step. Additionally, \( V_{sf,b} \) – processing volume for seed fermentation (L), \( V_{sf,d} \) – smallest denomination of seed fermenter volume available, \( V_{sf} \) – commercial size of fermenter volume required, and \( q \) – space efficiency factor taken as 0.75 which is used ubiquitously across the mathematical model. \( V_{sf,max} \) and \( V_{sf,d} \) are respectively taken as 20,000L and 1,000L.

Mass balance equations:

\[
m_{i,sf,media} = m_{i,sf} - m_{i,sf} = \sum_x V_{sf,x,b} \rho_x - m_{i,sf} \quad (B.12)
\]

where \( m_{i,sf,media} \) – mass of media inflow to be added to the seed fermentation operation (kg), \( V_{sf,x,b} \) – volume of stream component \( x \) in the process volume of the seed fermenter, \( \rho_x \) – density of stream component \( x \), and \( m_{i,sf} \) – total mass outflow to fermenter (kg). At this stage, the densities of stream components are all assumed to be 1kg/L.

Cost relationships:

\[
E_{EQP,sf} = E_{sf} n_{sf} \quad (B.13)
\]
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

\[ E_{\text{DRM,sf}} = E_{\text{medial,sf}} n_{sf} V_{sf} + V_{sf} n_{sf} (3E_{\text{CIP/L}} + 3E_{\text{WFI/L}}) \]  
(B.14)

\[ E_{\text{DU,sf}} = (7.5E_{\text{steam/L}} + 2.5E_{\text{cw/L}}) V_{sf} n_{sf} \]  
(B.15)

\[ E_{\text{LBR,sf}} = E_{\text{staff/hr}} n_{\text{staff,sf}} t_{sf} \]  
(B.16)

\[ n_{\text{staff,inc}} = \begin{cases} 
2 & \text{if } n_{\text{flask}} \leq 3 \\
3 & \text{otherwise} 
\end{cases} \]  
(B.17)

\[ E_{\text{OPR,inc}} = E_{\text{LBR,inc}} + E_{\text{DRM,inc}} \]  
(B.18)

where \( E_{\text{EQP,sf}} \) – total equipment cost, \( E_{\text{DRM,sf}} \) – direct raw materials cost, \( E_{\text{DU,sf}} \) – direct utilities cost, \( E_{\text{CIP/L,sf}} \) – clean-in-place buffer cost per litre, \( E_{\text{WFI/L,sf}} \) – water for injection cost per litre, \( E_{\text{steam/L}} \) – steam cost per litre, \( E_{\text{cw/L}} \) – cooling water cost per litre, \( E_{\text{LBR,sf}} \) – total labour cost, \( E_{\text{staff/hr}} \) – cost of technical staff per hour, \( n_{\text{staff}} \) – number of technical staff assigned, and \( t_{sf} \) – total processing time. For whole monoclonal antibody manufacture \( t_{sf} \) is taken as 346 hours and for antibody fragment manufacture \( t_{sf} \) is taken as 144, which includes the time taken for the necessary cleaning cycles.

B.3 FERMENTATION

In fermentation substrates such as media undergo biochemical reactions, resulting in cell growth and product formation, which in the cases considered are whole monoclonal antibodies and monoclonal antibody fragments. The whole manufacture of monoclonal antibodies usually relies on a mammalian culture expression system such as Chinese hamster ovary (CHO) cells. The manufacture of antibody fragments is achieved by bacterial fermentation in organisms such as \( E. \ coli \).

The equipment sizing equations are:

\[ V_{f,b} = \frac{D}{TYn_{b} p_{b}} \]  
(B.19)

\[ n_{f} = \left[ \frac{V_{f,b}}{qV_{f,\text{max}}n_{f}} \right] \]  
(B.20)

\[ V_{f} = \left[ \frac{D}{TYn_{b} p_{b} V_{f,d} n_{f} q} \right] V_{f,d} \]  
(B.21)
where the subscript \( f \) denotes reference to fermentation, \( V_{fb} \) – volume of fermentation broth used (L), \( D \) – annual output of the facility (kg), \( T \) – product titre in the fermentation broth (kg/L), \( Y \) – product yield achieved after downstream processing, \( n_B \) – number of batch cycles that take place in one year, \( p_B \) - probability that a batch is run successfully, \( n_f \) – number of fermentation units required, \( V_{f,max} \) – maximum volume of fermenter available, \( V_{fd} \) – smallest denomination of fermenter volume available, and \( V_f \) commercial size of fermenter volume required.

The mass inflow equations are as follows:

\[
m_{i,f,\text{media}} = m_{o,f} - m_{i,f} = \sum_x V_{f,xb} \rho_x - m_{i,f} \tag{B.22}
\]

\[
m_{i,f,\text{cells}} = m_{o,f,\text{cells}} \tag{B.23}
\]

\[
m_{i,f,\text{prod}} = 0 \tag{B.24}
\]

where \( m_{i,f,\text{media}} \) – media mass inflow to fermenter (kg), \( m_{o,f} \) – mass outflow from fermenter, \( m_{i,f} \) – mass inflow to fermenter, \( m_{i,f,\text{cells}} \) – cell mass inflow to fermenter (kg), \( V_{f,xb} \) – volume of stream component \( x \) in the process volume of the fermenter, \( m_{i,f,\text{prod}} \) – product mass inflow to fermenter, \( \rho_f \) – density of fermentation broth.

The mass outflow equations are:

\[
m_{o,f,\text{media}} = m_{i,f,\text{media}} (1 - r) \tag{B.25}
\]

\[
m_{o,f,\text{cells}} = m_{i,f,\text{cells}} - m_{i,f,\text{media}} r \frac{s_{\text{cells}}}{s_{\text{media}}} \tag{B.26}
\]

\[
r = -\frac{m_{o,f,\text{prod}} s_{\text{media}}}{m_{i,f,\text{media}} s_{\text{prod}}} \tag{B.27}
\]

where \( m_{o,f,\text{media}} \) – media mass outflow to fermenter (kg), \( m_{o,f,\text{cells}} \) – cell mass outflow to fermenter (kg), \( s \) is the stoichiometric ratio of the corresponding substance in the fermentation reaction and \( r \) is the reaction extent. It is assumed here that \( s_{\text{cells}} = 0.5 \), \( s_{\text{media}} = -1.5 \), and \( s_{\text{prod}} = 0.5 \). Here, mass stoichiometry and the extent of reaction were
used to describe the relationship between substrate consumption and product formation. Intuitively, the stoichiometric coefficients used in the model were negative for the media, which contains the substrate, and positive for product and cells.

The cost relationships are:

\[
E_{EQP,f} = (83011\ln(V_f) - 14889)n_f \quad \text{(B.28)}
\]

\[
E_{DRM,f} = E_{\text{media}} \cdot m_{i,f,\text{media}} + V_f n_f (3E_{CIP/L} + 3E_{\text{WFI/L}}) \quad \text{(B.29)}
\]

\[
E_{DU,f} = (7.5E_{\text{steam/L}} + 2.5E_{\text{ew/L}}) V_f n_f \quad \text{(B.30)}
\]

\[
E_{LBR,f} = E_{\text{staff/hr}} n_{\text{staff,f}} t_f \quad \text{(B.31)}
\]

\[
n_{\text{staff,f}} = \begin{cases} 
2 & n_f \leq 3 \\
3 & \text{otherwise}
\end{cases} \quad \text{(B.32)}
\]

\[
E_{OPR,f} = E_{EQP,f} + E_{DRM,f} + E_{LBR,f} \quad \text{(B.33)}
\]

where \( E_f \) – total equipment cost, \( E_{DRM,f} \) – direct raw materials cost, \( E_{DU,f} \) – direct utilities cost, \( E_{CIP/L,f} \) – clean-in-place buffer cost per litre, \( E_{\text{WFI/L,f}} \) – water for injection cost per litre, \( E_{\text{steam/L}} \) – steam cost per litre, \( E_{\text{ew/L}} \) – cooling water cost per litre, \( E_{LBR,f} \) – total labour cost, \( E_{\text{staff/hr}} \) – cost of technical staff per hour, \( n_{\text{staff}} \) – number of technical staff assigned, and \( t_f \) – total processing time for fermentation. For whole monoclonal antibody manufacture \( t_f \) is taken as 346 hours and for antibody fragment manufacture \( t_f \) is taken as 144, which includes the time taken for the necessary cleaning cycles. The cost relationship \( E_f \) is derived from a logarithmic regression fit for the price of fermenters over a range of volumes (Figure B.1). It was found that this type of regression was able to explain the fermenter size and price data most accurately. Using the geometric capital size and cost relationship seen in Equation 5.1 may very well be inappropriate in this case. This is because selecting an appropriate reference fermenter to make an estimate from is difficult. This is illustrated in Table B.2. It can be observed that using the 300L or 1000L fermenter as a reference results in gross overestimation of the 15000L fermenter. Also, use of the 15,000L fermenter as a reference results in significant underestimation of the 300L and 1000L fermenters. It can also be seen that using the logarithmic relationship derived from Figure B.1
results in the most accurate estimations for fermenter costs included in the dataset overall.

![Cost versus volume for a fermenter. A logarithmic regression fit to the data is also shown where $f(x) = 73254 \ln(x) - 64605$, and correlation coefficient is 0.9002. Data source: Personal discussion with C. Osborn at the ACBE, UCL (2004), and Jacobs (2006).](image)

**Figure B.1.** Cost versus volume for a fermenter. A logarithmic regression fit to the data is also shown where $f(x) = 73254 \ln(x) - 64605$, and correlation coefficient is 0.9002. Data source: Personal discussion with C. Osborn at the ACBE, UCL (2004), and Jacobs (2006).

**Table B.2.** Comparison of predictors for fermenter costs.

<table>
<thead>
<tr>
<th>Fermenter Size (L)</th>
<th>Actual</th>
<th>300L</th>
<th>1000L</th>
<th>1500L</th>
<th>73254ln(x) - 64605</th>
</tr>
</thead>
<tbody>
<tr>
<td>100L</td>
<td>£301282</td>
<td>£144839</td>
<td>£123082</td>
<td>£31166</td>
<td>£272742</td>
</tr>
<tr>
<td>300L</td>
<td>£280000</td>
<td>£280000</td>
<td>£237941</td>
<td>£60250</td>
<td>£353220</td>
</tr>
<tr>
<td>1000L</td>
<td>£490000</td>
<td>£576614</td>
<td>£490000</td>
<td>£124075</td>
<td>£441416</td>
</tr>
<tr>
<td>2000L</td>
<td>£498066</td>
<td>£873984</td>
<td>£742701</td>
<td>£188063</td>
<td>£492192</td>
</tr>
<tr>
<td>15000L</td>
<td>£630000</td>
<td>£2927791</td>
<td>£2488003</td>
<td>£630000</td>
<td>£639791</td>
</tr>
</tbody>
</table>

*Note: The column labelled 'Actual' denotes the actual price of each specified fermenter. Columns labelled '300L', '1000L', and '1500L' refer to the estimated fermenter costs when these fermenter sizes are used as the base reference, or $\text{Size}_i$ and $\text{Cost}_i$, in equation 5.1. The column labelled '73254ln(x) - 64605' denotes to the estimated fermenter costs when this relationship is used. Data source: Personal discussion with C. Osborn at the ACBE, UCL (2004).*

### B.4 Ultrafiltration

It is assumed that for the intended range of inputs that the use of one ultrafiltration unit is sufficient.

Relationships for the mass balance are:
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

\[ m_{\text{ret},x,uf} = m_{i,x,uf} F_c (G_{x}-1) \]  
(B.34)

\[ C_{\text{r,prod}} = \frac{\ln(Y_{uf})}{\ln(F_c)} + 1 \]  
(B.35)

\[ m_{\text{perm},x,uf} = m_{i,x,uf} - m_{\text{ret},x,uf} \]  
(B.36)

\[ Y_{uf} = Y^{1/n_{DISP}} \]  
(B.37)

where the subscript \( uf \) denotes belonging to the ultrafiltration unit, \( m_{\text{ret},x} \) – mass of stream component \( x \) in the retentate, \( m_{i,x,uf} \) – the mass inflow of stream component \( x \), \( F_c \) – the concentration factor, \( C_{\text{r},x} \) – rejection coefficient for stream component \( x \), and \( Y_{uf} \) – yield of the ultrafiltration unit, \( n_{DISP} \) – number of downstream processing units have non-negligible effects on the mass of product in the stream. The values for \( C_{\text{r,media}}, C_{\text{r,cells}}, \) and \( C_{\text{r,buffer}} \) are respectively taken as 0, 1, and 0.

Relationships for processing time are:

\[ t_{\text{cycle},uf} = \frac{V_{i,uf} \left(1 - \frac{1}{F_c}\right)}{J_{uf} A_{uf} n_{uf}} \]  
(B.38)

\[ t_{uf} = (t_{CIP,uf} + t_{\text{cycle},uf}) n_{\text{cycle},uf} \]  
(B.39)

where \( t_{\text{cycle}} \) – cycle time, \( V_{i,uf} \) – volume inflow to the ultrafiltration unit, \( J_{uf} \) – average flux taken to be 100L/m²/hr, \( A_{uf} \) – total membrane area, \( n_{uf} \) – number of ultrafiltration units, \( t_{CIP,uf} \) – CIP completion time, and \( n_{\text{cycle},uf} \) – number of cycles.

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrafiltration rig</td>
<td>£37,500</td>
</tr>
<tr>
<td>Filter cartridge, ( A_{uf} = 0.93m^2 )</td>
<td>£883</td>
</tr>
<tr>
<td>Filter cartridge, ( A_{uf} = 1.9m^2 )</td>
<td>£1763</td>
</tr>
<tr>
<td>Filter cartridge, ( A_{uf} = 3.7m^2 )</td>
<td>£3228</td>
</tr>
</tbody>
</table>

Source: Millipore.
The prices of required holding tanks were based on a linear regression fit to prices of available data (Table B.3), which produced the relationship:

\[ E_{ht} = 72549 \left[ \frac{V_{ret,uf}}{qV_{ht,d}} \right] V_{ht,d} + 1137.3 \]  

(B.40)

where the subscript \( ht \) denotes reference to the holding tank, \( V_{ret,uf} \) – volume of the retentate, and \( V_{ht,d} \) – smallest denomination of holding tank size available. The related cost relationships are:

\[ E_{EQP,uf} = E_{ht,ret,uf} + E_{rig,uf} \]  

(B.41)

\[ E_{DRM,uf} = V_{ht,ret,uf} n_{ht,ret,uf} (E_{CIP/L} + E_{WFI/L}) \]  

(B.42)

\[ E_{DU,uf} = E_{filter,uf} n_{filter,uf} = E_{filter,uf} \frac{V_{uf}}{r_{filter,uf}} \]  

(B.43)

\[ E_{LBR,uf} = E_{staff/hr} n_{staff,uf} t_{uf} \]  

(B.44)

\[ n_{staff,uf} = 2 \]  

(B.45)

\[ E_{OPR,uf} = E_{DU,uf} + E_{DRM,uf} + E_{LBR,uf} \]  

(B.46)

where \( E_{EQP,uf} \) – total equipment cost for ultrafiltration, \( E_{ht,ret,uf} \) – holding tank cost for retentate, \( E_{rig,uf} \) – cost of one ultrafiltration rig, \( V_{ht,ret,uf} \) – volume of retentate holding tank, and \( r_{filter,uf} \) – rate at which the ultrafiltration filter cartridge is replaced. The subscript \( filter \) denotes reference to the ultrafiltration filter cartridge. The model selects the appropriate size of filter based on the volume throughput requirements for the ultrafiltration process which are referenced to the costs exhibited in Table B.3.

**B.5 CHROMATOGRAPHY**

Chromatography is used not only to eliminate contaminants but to also capture products. Chromatography techniques allow for the binding specificity of the chromatography matrix to the product to separate it from the crude product stream. Broadly, the operation involves the loading of the inlet stream to the chromatography column, the washing of the matrix to remove unbound materials, usually
contaminants, followed by an elution step to recover the bound product. The matrix is then re-equilibrated in preparation for the next cycle. The types of chromatography considered between the two manufacturing processes are affinity and ion-exchange chromatography.

The equipment sizing relationships for the chromatography column are as follows:

\[ V_{i,chr} = \frac{m_{i,prod,chr}}{c_b n_{cycle,chr}} \]  \hspace{1cm} (B.47)
\[ d_{chr} = \left[ \frac{1}{d_{chr,d}} \sqrt{\frac{4V_{i,chr}}{n_{chr} n_{chr}}} \right] d_{chr,d} \]  \hspace{1cm} (B.48)
\[ n_{chr} = \left[ \frac{d_{chr}}{d_{chr,\text{max}}} \right] \]  \hspace{1cm} (B.49)
\[ A_{chr} = 0.25 \pi d_{chr}^2 \]  \hspace{1cm} (B.50)
\[ V_{chr} = h_{chr} A_{chr} \]  \hspace{1cm} (B.51)
\[ t_{chr} = \frac{n_{cycle,chr}}{Q_{chr}} \sum_{j} V_{chr} v_j \]  \hspace{1cm} (B.52)

where the subscript \( chr \) denotes reference to the chromatographic operation, \( c_b \) – binding capacity of the chromatography matrix (kg/L), \( d_{chr} \) – chromatography column diameter, \( d_{chr,d} \) – smallest denomination of column diameter, \( h_{chr} \) – chromatography column height, \( d_{chr,\text{max}} \) – maximum available column diameter, \( n_{chr} \) – number of chromatography columns, \( A_{chr} \) – cross-sectional area of column, \( V_{chr} \) – column volume, and \( v_j \) – number of column volumes of buffer required for each \( j^{th} \) step in the operation, as shown in Table B.4.
Table B.4. Column volumes, \( v_i \), required for affinity chromatography operations used.

<table>
<thead>
<tr>
<th>Step</th>
<th>( v_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Equilibration A</td>
<td>5</td>
</tr>
<tr>
<td>2. Equilibration B</td>
<td>5</td>
</tr>
<tr>
<td>3. Post load wash</td>
<td>5</td>
</tr>
<tr>
<td>4. Elution</td>
<td>2</td>
</tr>
<tr>
<td>5. Regeneration A</td>
<td>5</td>
</tr>
<tr>
<td>6. Regeneration B</td>
<td>5</td>
</tr>
<tr>
<td>7. Cleaning</td>
<td>5</td>
</tr>
<tr>
<td>8. Sanitisation</td>
<td>10</td>
</tr>
<tr>
<td>9. Storage</td>
<td>3</td>
</tr>
</tbody>
</table>

Table B.5. Column volumes, \( v_i \), required for anion and cation exchange chromatography operations.

<table>
<thead>
<tr>
<th>Step</th>
<th>( v_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Equilibration</td>
<td>5</td>
</tr>
<tr>
<td>2. Post load wash 1</td>
<td>3</td>
</tr>
<tr>
<td>3. Post load wash 2</td>
<td>5</td>
</tr>
<tr>
<td>4. Elution</td>
<td>10</td>
</tr>
<tr>
<td>5. Wash</td>
<td>5</td>
</tr>
<tr>
<td>6. Reequilibration</td>
<td>5</td>
</tr>
<tr>
<td>7. Cleaning</td>
<td>5</td>
</tr>
<tr>
<td>8. Sanitisation</td>
<td>10</td>
</tr>
<tr>
<td>9. Storage</td>
<td>3</td>
</tr>
</tbody>
</table>

Mass balance equations:

\[
m_{o, \text{prod,chr}} = m_{i, \text{prod,chr}} y^{1/n_{\text{step}}} \quad (B.53)
\]

\[
m_{o, \text{media,chr}} = 0 \quad (B.54)
\]

\[
m_{o, \text{cells,chr}} = 0 \quad (B.55)
\]

\[
m_{o, hfr, chr} = 3V_{chr} n_{\text{cycle}} \quad (B.56)
\]

where the subscript \( hfr \) refers to storage buffer.
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

Cost relationships:

\[ E_{ht.chr.j} = 72549 \left( \frac{V_{chr.j} n_{cycle.chr}}{q V_{ht,d}} \right) V_{ht,d} + 1137.3 \]  \hspace{1cm} (B.57)

\[ E_{EQP.chr} = n_{chr} (E_{rig.chr} + E_{chr}) + \sum_j E_{ht.chr.j} \]  \hspace{1cm} (B.58)

\[ E_{DRM.chr} = \frac{V_{chr} n_{cycle.chr}}{r_{ mtx}} E_{mtx/L} + \sum_j E_{buffer/L,j} n_{chr} V_j n_{cycle.chr} \]  \hspace{1cm} (B.59)

\[ E_{DU.chr} = 0 \]  \hspace{1cm} (B.60)

\[ E_{LBR.chr} = E_{staff/hr} n_{staff.chr} t_{chr} \]  \hspace{1cm} (B.61)

\[ n_{staff.chr} = \begin{cases} 2 & n_{chr} \leq 3 \\ 3 & \text{otherwise} \end{cases} \]  \hspace{1cm} (B.62)

\[ E_{OPR.chr} = E_{LBR.chr} + E_{DRM.chr} \]  \hspace{1cm} (B.63)

where \( E_{ht.chr.j} \) - holding tank expense for the buffer of the \( j^{th} \) chromatography step, \( E_{rig.chr} \) - cost the chromatography rig, \( E_{chr} \) - cost of the chromatography column, \( r_{ mtx} \) - reusability of the chromatography matrix represented as the number of wash cycles it can be used for before it must be discarded, \( E_{mtx/L} \) - cost chromatography matrix per litre, and \( E_{buffer} \) - cost of buffer per litre. It should be noted that when the same buffers are being used for different chromatography step then a single holding tank can house the corresponding buffer. The values of \( r_{ mtx} \) used for Protein-A and ion-exchange media are respectively 300 and 250. The costs for matrices, equipment, and buffers used are given in Table B.6 and Table B.7.

<table>
<thead>
<tr>
<th>Table B.6. Chromatography matrix costs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein-A Matrix</td>
</tr>
<tr>
<td>Ion Exchange Matrix</td>
</tr>
</tbody>
</table>
### Table B.7. Chromatography equipment costs.

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromatography column, diameter= 0.7 dm</td>
<td>£4,947</td>
</tr>
<tr>
<td>Chromatography column, diameter= 1.0 dm</td>
<td>£5,686</td>
</tr>
<tr>
<td>Chromatography column, diameter= 1.4 dm</td>
<td>£7,740</td>
</tr>
<tr>
<td>Chromatography column, diameter= 2.0 dm</td>
<td>£8,529</td>
</tr>
<tr>
<td>Chromatography column, diameter= 2.5 dm</td>
<td>£14,065</td>
</tr>
<tr>
<td>Chromatography column, diameter= 3.0 dm</td>
<td>£18,591</td>
</tr>
<tr>
<td>Chromatography column, diameter= 3.5 dm</td>
<td>£25,372</td>
</tr>
<tr>
<td>Chromatography column, diameter= 4.5 dm</td>
<td>£46,332</td>
</tr>
<tr>
<td>Chromatography Rig</td>
<td>£140,000</td>
</tr>
</tbody>
</table>

*Note: All chromatography columns represented have a bed height of 1.1 dm. Source: Millipore.*

### Table B.8. Chromatography buffer costs.

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0M NaOH</td>
<td>£1.0807/L</td>
</tr>
<tr>
<td>1.0M NaCl</td>
<td>£0.7770/L</td>
</tr>
<tr>
<td>1M Glycine/NaOH + 0.30 NaCl, pH8.6</td>
<td>£2.1390/L</td>
</tr>
<tr>
<td>1M Glycine/NaOH + 0.15 NaCl, pH8.6</td>
<td>£2.0972/L</td>
</tr>
<tr>
<td>0.1M Borate/0.15 NaCl, pH8.5</td>
<td>£0.5691/L</td>
</tr>
<tr>
<td>Phosphate Buffered Saline</td>
<td>£0.6570/L</td>
</tr>
<tr>
<td>0.1M Citrate, pH6.0-3.0</td>
<td>£0.7358/L</td>
</tr>
<tr>
<td>0.1M Glycine/HCl, pH3.5-3.0</td>
<td>£0.6556/L</td>
</tr>
<tr>
<td>HCl pH 1.5</td>
<td>£0.5243/L</td>
</tr>
<tr>
<td>H₃PO₄ pH 1.5</td>
<td>£0.6301/L</td>
</tr>
<tr>
<td>6M Guanidine HCl</td>
<td>£9.3105/L</td>
</tr>
<tr>
<td>20% Ethanol</td>
<td>£0.6408/L</td>
</tr>
<tr>
<td>0.01M Tris/HCl, pH 7.0</td>
<td>£0.5505/L</td>
</tr>
<tr>
<td>0.001M Glycine/6M Urea, pH 4.55</td>
<td>£3.4375/L</td>
</tr>
<tr>
<td>0.01M Tris/HCl with 0.01M NaCl</td>
<td>£0.5532/L</td>
</tr>
<tr>
<td>0.02M Sodium Citrate/0.1M NaCl, pH 7.0</td>
<td>£0.6075/L</td>
</tr>
</tbody>
</table>

*Source: Adapted from Fisher Scientific. Note: The above substances are normally sold in powdered form for which the original prices are obtained. The costs reflect the resultant cost when converted to the liquid formulation required for processing operations. These can be conveniently used in the model as usage is based on the volume of the respective vessel.*
Figure B.2. Cost versus volume relationship for holding tanks. The solid line represents the line of best fit as determined by geometric regression. The equation of this line is \( f(x) = 244.33x^{0.6087} \), with \( R^2 = 0.9988 \). Data source: Personal discussion with C. Osborn at the ACBE, UCL (2004), and Jacobs (2006).

The cost relationship of the holding tank relative to volume is displayed in Figure B.2. The \( R^2 \) value is very high demonstrating that a large proportion of the variance in cost can be explained by the derived equation.

B.6 VIRAL INACTIVATION

Viral inactivation is the first viral clearance step in the process. The incoming crude product stream is added to a formulation of acid designed to inactivate viral molecules. As this is not a separation process it is assumed that the product loss is negligible and that the yield for the process is equal to 1.

The mass balance relationships are straight forward:

\[
m_{o,vi,x} = m_{i,vi,x} \quad (B.64)
\]

\[
m_{i,vi,acid} = 0.01 \sum_x m_{i,vi,x} \rho_x / \rho_{acid} \quad (B.65)
\]

where the subscript \( vi \) indicates reference to the viral inactivation unit operation, \( m_{o,vi,x} \) - mass outflow of stream component \( x \) (kg), and \( m_{i,vi,x} \) - mass inflow of stream
component $x$ (kg). The outflow stream also includes the acid added during the procedure. It should be noted that it is taken that the volume of acid added is at a ratio of 1:100 with the incoming process stream volume.

The processing time is represented as:

$$t_{vi} = (t_{CIP, vi} + t_{SIP, vi} + t_{cycle, vi}) n_{cycle, vi}$$

(B.66)

where $t_{vi}$ – total processing time, $t_{CIP, vi}$ – time taken for clean-in-place, and $t_{SIP, vi}$ – time taken for steam-in-place. The values $t_{CIP, vi}$, $t_{SIP, vi}$, $t_{cycle, vi}$, and $n_{cycle, vi}$ are taken as 2 hours, 6 hours, 4 hours, and 1 hour respectively.

The cost relationships are:

$$E_{h, vi} = 72549 \left[ \frac{V_{i, vi} n_{cycle, vi}}{qV_{h, d}} \right] V_{h, d} + 1137.3$$

(B.67)

$$V_{h, vi} = \left[ \frac{V_{i, vi} n_{cycle, vi}}{qV_{h, d}} \right] V_{h, d}$$

(B.68)

$$E_{EQP, vi} = E_{h, vi}$$

(B.69)

$$E_{DRM, vi} = E_{acid/L} V_{i, vi, acid} n_{cycle, vi}$$

(B.70)

$$E_{DU, vi} = (7.5 E_{stream/L} + 2.5 E_{CIP/L} + 2.5 E_{WFI/L}) V_{h, vi}$$

(B.71)

$$E_{LBR, vi} = E_{staff/\text{hr}} n_{staff, vi} t_{vi}$$

(B.72)

$$n_{staff, \text{chr}} = 1$$

(B.73)

$$E_{OPR, vi} = E_{LBR, vi} + E_{DRM, vi}$$

(B.74)

where $E_{acid/L}$ – cost of viral inactivation acid per litre.

**B.7 Filtration Operations**

This section covers the diafiltration, viral nanofiltration, and dead-end filtration operations. Such filtration operations are characterised by exploiting the differences
in particle sizes to separate particles found in the crude product stream by passing the stream through a filtration membrane. The portion of the stream retained by the filter membrane is defined as the retentate and the particles that passes through the membrane are defined as the permeate. Filtration operations are often used to recover products and to concentrate the product stream. Viral nanofiltration has the added function of assisting in the sterilisation of the product stream through design of the membrane filter which removes viral particles. Cross-flow filtration employs tangential flow across the membrane surface to reduce fouling of the membrane.

Mass balance relationships are:

\[
\begin{align*}
    m_{\text{ret,s,sf}} &= m_{\text{i,s,sf}} e^{u(1-C_{r,s})} \\
    C_{r,\text{prod}} &= \frac{\ln(Y_{sf})}{\ln(u)} + 1 \\
    m_{\text{perm,s,sf}} &= m_{\text{i,s,sf}} - m_{\text{ret,s,sf}} \\
    Y_{sf} &= Y^{1/u_{\text{var}}}
\end{align*}
\]

where the subscript \(sf\) denotes referral to diafiltration, virus nanofiltration, and dead-end filtration processes, \(u\) – number of filtration volumes used which is taken as 6. The values of \(C_{r,s}\) used are exhibited in Table B.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>Viral Nanofiltration</th>
<th>Diafiltration</th>
<th>Dead-End Filtration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Media</td>
<td>0.99</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cells</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Buffer</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Acid</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>


Process time relationships are as follows:

\[
t_{sf} = (t_{\text{CIP,sf}} + t_{\text{SIP,sf}} + t_{\text{cycle,sf}}) n_{\text{cycle,sf}}
\]
\[ t_{\text{cycle,sf}} = \frac{V_{\text{i,sf}} u}{J A_{\text{sf}} n_{\text{sf}}} \]  
(B.80)

\[ A_{\text{sf}} = \frac{V_{\text{i,sf}} u}{J_{\text{sf}} n_{\text{sf}}} \]  
(B.81)

where \( t_{\text{cycle,sf}} \) - cycle time, \( V_{\text{i,sf}} \) - volume inflow to the filtration unit, \( J_{\text{sf}} \) - average flux taken to be 100L/m²/hr, \( A \) - total membrane area, \( n_{\text{sf}} \) - number of ultrafiltration units, \( t_{\text{CIP,sf}} \) - CIP completion time, \( t_{\text{SIP,sf}} \) - SIP completion time, and \( n_{\text{cycle,sf}} \) - number of processing cycles.

### Table B.10. Relevant equipment price data for the ultrafiltration process.

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtration rig</td>
<td>£192,373</td>
<td>Jacobs</td>
</tr>
<tr>
<td>Filter cartridge, viral inactivation, ( A_{\text{sf}} = 0.48m^2 )</td>
<td>£2,142</td>
<td>Millipore</td>
</tr>
<tr>
<td>Filter cartridge, viral inactivation, ( A_{\text{sf}} = 0.70m^2 )</td>
<td>£5,091</td>
<td>Millipore</td>
</tr>
<tr>
<td>Filter cartridge, viral inactivation, ( A_{\text{sf}} = 1.4m^2 )</td>
<td>£10,182</td>
<td>Millipore</td>
</tr>
<tr>
<td>Filtration buffer</td>
<td>£1.50/L</td>
<td>Millipore</td>
</tr>
</tbody>
</table>

Cost relationships are as follows:

\[ E_{\text{w,sf}} = 72549 \left[ V_{\text{ret,sf}} \right] \frac{V_{\text{het,d}}}{q V_{\text{het,d}}} + 1137.3 \]  
(B.82)

\[ E_{\text{eqp,sf}} = E_{\text{het,ret,sf}} + E_{\text{rig,sf}} \]  
(B.83)

\[ E_{\text{Drm,sf}} = V_{\text{het,ret,sf}} n_{\text{het,ret,sf}} (E_{\text{CIP/L}} + E_{\text{SIP/L}} + E_{\text{WFI/L}}) + V_{\text{buff,sf}} E_{\text{buff/L}} \]  
(B.84)

\[ V_{\text{buff,sf}} = V_{\text{perm,sf}} \]  
(B.85)

\[ E_{\text{Duf,sf}} = E_{\text{filter,sf}} n_{\text{filter,sf}} = E_{\text{filter,sf}} \frac{V_{\text{i,sf}}}{r_{\text{filter,sf}}} \]  
(B.86)

\[ E_{\text{LBR,sf}} = E_{\text{staff/hr}} n_{\text{staff,sf}} \]  
(B.87)

\[ n_{\text{staff,sf}} = 1 \]  
(B.88)

\[ E_{\text{opr,sf}} = E_{\text{LBR,sf}} + E_{\text{Drm,sf}} \]  
(B.89)
where $E_{EQP,xf}$ - total equipment cost for filtration, $E_{ht,ret,xf}$ - holding tank cost for retentate, $E_{rig,xf}$ - cost of one filtration rig, $V_{ht,ret,xf}$ - volume of retentate holding tank, $V_{buff,xf}$ - volume of filtration buffer required, $E_{buff/l}$ - cost of filtration buffer per litre, $V_{perm,xf}$ - volume of permeate stream, and $r_{filter,xf}$ - rate at which the filtration filter cartridge is replaced. The model selects the appropriate size of filter based on the volume throughput requirements for the ultrafiltration process which are referenced to the costs exhibited in Table B.10. It should be noted that in the model the viral nanofiltration and dead-end filters utilise the same model of filter cartridge, and that diafiltration uses no filter cartridges.

### B.8 CENTRIFUGATION

Centrifugation is used to separate out the solids from the liquids from a mixture, based on the components’ size and density. As seen in Equation 5.1, it is used in as a primary separation operation to remove cells and cell debris via the application of centrifugal forces to produce the sediment and the supernatant. The output stream was calculated using the dewatering level that can be achieved and the solid carry-over level. The model assumes that there was a single component in the solid phase, cells. The sediment stream and the supernatant stream were calculated separately.

Mass balance relationships:

\[ m_{ctr,sp,cells} = m_{i,cells,ctr} C_s \]  
\[ m_{ctr,sd,cells} = m_{i,cells,ctr} - m_{cells,sp} \]  
\[ m_{ctr,sd,media} = \left( \frac{m_{ctr,sd,cells}}{\rho_{cells}} - \frac{m_{ctr,sd,cells}}{\rho_{cells}} \right) \rho_{media} \]  
\[ L_d = \frac{m_{ctr,sd,cells}}{(1 - Y^{1/DSP})m_{i,cells} + m_{ctr,sd,cells}} \]  
\[ m_{ctr,sp,media} = m_{i,media} - m_{ctr,sd,media} \]  
\[ m_{ctr,sp,product} = \frac{m_{i,product} m_{ctr,sp,media}}{m_{i,media}} \]  
\[ m_{ctr,sp,media} = m_{i,media} - m_{ctr,sd,media} \]
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

Where the subscript $ctr$ specifies reference to the centrifugation process, $m_{ctr,sp,cell}$ - cell mass in the supernatant stream, $m_{i,cell,ctr}$ - inflowing mass of the cells to the centrifuge, $C_s$ - solid carry over level which is taken here as 0.02, $m_{ctr,sd,cell}$ - cell mass in the sediment stream, $m_{ctr,media}$ - media mass in the sediment stream, $L_d$ - dewatering level, and $m_{sd,media}$ - media mass in the sediment stream.

Processing time relationships:

$$t_{ctr} = n_{cycle,ctr} (t_{cycle,ctr} + t_{CIP,ctr})$$ (B.97)

$$t_{cycle,ctr} = \frac{V_{i,ctr}}{Q_{ctr}}$$ (B.98)

where $t_{ctr}$ - total processing time taken for the centrifugation process, $n_{cycle,ctr}$ - number of cycles for which the operation is run which is taken as 1, $t_{cycle,ctr}$ - throughput time for the process stream, $t_{CIP,ctr}$ - CIP time taken as 6 hours, $V_{i,ctr}$ - inflow volume of process stream to the centrifuge, and $Q_{ctr}$ - operating flowrate of centrifuge.

The prices of required holding tanks were based on a linear regression fit to prices of available data (Table B.3), which produced the relationship:

$$E_{ht,ctr,sp} = 72549 \left[ \frac{V_{ht,ctr,sp}}{qV_{ht,d}} \right] V_{ht,d} + 1137.3$$ (B.99)

$$E_{EQP,ctr} = E_{ctr} n_{ctr} + E_{ht,ctr,sp}$$ (B.100)

$$n_{ctr} = \left[ \frac{t_{cycle,ctr}}{6} \right]$$ (B.101)

$$E_{DRM,ctr} = Q_{ctr} t_{cycle,ctr} \left( E_{CIP/L} + 2E_{SIP/L} \right)$$ (B.102)

$$E_{DU,ctr} = 2Q_{ctr} t_{cycle,ctr} \left( E_{steam/L} + E_{cw/L} \right)$$ (B.103)

$$E_{LBR,ctr} = E_{staff/hr} n_{staff,ctr} t_{ctr}$$ (B.104)

236
APPENDIX B: MODELLING EQUATIONS USED IN CHAPTER 5

\[ n_{\text{staff, ctr}} = \begin{cases} 1 & n_{\text{ctr}} \leq 3 \\ 2 & \text{otherwise} \end{cases} \]  
(B.105)

\[ E_{\text{OPR, ctr}} = E_{\text{LBR, ctr}} + E_{\text{DRM, ctr}} \]  
(B.106)

where \( E_{\text{ht, ctr, sp}} \) - cost of the holding tank for the supernatant, \( E_{\text{EQP, ctr}} \) - total equipment cost for centrifugation, \( E_{\text{ctr}} \) - cost of one centrifuge, and \( n_{\text{ctr}} \) - number of centrifuges required. The centrifuge used for the process is assumed to have a flowrate of 1500L/hr and costs £474,995 (Source: Jacobs).

**B.9 Periplasmic extraction**

The monoclonal antibody fragments produced by the *E.coli* cells are exported to the periplasmic space. Periplasmic extraction enables gentle and efficient solubilisation of the outer membrane of *E.coli* cells, permitting rapid extraction of the fragments from the periplasmic space. This can result in an increase in product yield compared to other extraction methods such as homogenisation and lysis of the entire cell (Cossins *et al.* 2007). The yield for this unit operation was assumed to be 100%, as the unit operation does not involve any types of separation of the input stream.

Equipment sizing and process time relationships:

\[ n_{\text{ht, ppe}} = \left[ \frac{V_{i, ppe}}{qV_{\text{ht, max}}} \right] \]  
(B.107)

\[ V_{\text{ht, ppe}} = \left[ \frac{V_{i, ppe}}{qV_{\text{ht, d}}} \right] V_{\text{hi, d}} \]  
(B.108)

\[ t_{\text{ppe}} = n_{\text{cycle, ppe}} (t_{\text{cycle, ppe}} + t_{\text{CIP, ppe}} + t_{\text{SIP, ppe}}) \]  
(B.109)

where the subscript *ppe* denotes reference to the periplasmic extraction operation. The values \( n_{\text{cycle, ppe}}, t_{\text{cycle, ppe}}, t_{\text{CIP, ppe}} \), and \( t_{\text{SIP, ppe}} \) are taken as 1, 8, 6, and 4 respectively.

The mass balances are straightforward as this process is assumed to negligible loss in product.
Cost relationships:

\[ E_{\text{ht,cr}.\text{ppe}} = 72549 \left( \frac{V_{\text{ht,pppe}}}{qV_{\text{ht,d}}} \right) V_{\text{ht,d}} + 1137.3 \]  

(B.110)

\[ E_{\text{EQP}.\text{ppe}} = E_{\text{ht,pppe}} n_{\text{ht,pppe}} \]  

(B.111)

\[ E_{\text{DRM}.\text{pppe}} = n_{\text{ht,pppe}} V_{\text{ht,pppe}} (0.05E_{\text{TEX/L}} + E_{\text{CIP/L}} + E_{\text{WFI/L}}) \]  

(B.112)

\[ E_{\text{DU}.\text{pppe}} = n_{\text{ht,pppe}} V_{\text{ht,pppe}} (E_{\text{steam/L}} + E_{\text{cw/L}}) \]  

(B.113)

\[ E_{\text{LBR}.\text{pppe}} = E_{\text{staff/L,pppe,nstaff,pppe}} \]  

(B.114)

\[ n_{\text{staff,pppe}} = \begin{cases} 1 & n_{\text{ht,pppe}} \leq 3 \\ 2 & \text{otherwise} \end{cases} \]  

(B.115)

\[ E_{\text{OPR}.\text{pppe}} = E_{\text{LBR,pppe}} + E_{\text{DRM,pppe}} \]  

(B.116)

where \( E_{\text{TEX/L}} \) – cost of TEX buffer used taken as £1.50/L.

### B.10 pH AND CONDUCTIVITY ADJUSTMENT

A pH and conductivity adjustment is carried out after the periplasmic extraction so as to precipitate the product. The equipment sizing, mass balances and costs are almost identical to periplasmic extraction. The only exception is that \( t_{\text{cycle,pc}} = 2 \) hours, with \( pc \) denoting reference to this operation. Also no TEX buffer is used in this operation; instead an acid is used in a volumetric ratio of 1:200 to the process volume at a cost £0.50/L.

### B.11 DILUTION

In the dilution step, the product stream was diluted to a predefined concentration as a preparation for the following chromatography operation. Chromatography operations are sensitive and expensive thus such precautionary steps are required. As with pH adjustment and conductivity, the equipment sizing, mass balances and costs are almost identical to periplasmic extraction. The only exception is that no TEX buffer
is used in this operation; instead another buffer used in a volumetric ratio of 1:10 to the process volume at the same cost of £1.50/L.

**B.12 PEGylation**

PEGylation is applied to intravenously administered therapeutic proteins, in particular, to improve their stability, biological half-life, water solubility, and immunologic characteristics (Harris, 2003). The equipment sizing, mass balances and costs are almost identical to periplasmic extraction. The only exception is that $t_{cycle,PEG} = 12$ hours, with $PEG$ denoting reference to this operation. Also no TEX buffer is used in this operation; instead another buffer used in a volumetric ratio of 1:5 to the process volume at the same cost of £1.50/L.
APPENDIX C: PAPERS BY THE AUTHOR

The papers included in this Appendix are listed below:


A multi-criteria decision-making framework for the selection of strategies for acquiring biopharmaceutical manufacturing capacity

Edmund George, Nigel J. Titchener-Hooker, Suzanne S. Farid*

The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom

Received 5 December 2005; received in revised form 13 December 2006; accepted 15 December 2006

Available online 27 December 2006

Abstract

Selecting an appropriate path for acquiring commercial-scale biopharmaceutical manufacturing capacity often requires the rationalisation of multiple conflicting criteria and the reconciling of financial and non-financial issues. Multi-criteria decision-making (MCDM) can provide a holistic framework for evaluating such scenarios. This paper presents the development of a decision-support framework for decision-making scenarios involving the acquisition of commercial-scale biopharmaceutical manufacturing capacity that utilises MCDM. To illustrate the functionality of the framework, a hypothetical scenario was constructed based on a biopharmaceutical company faced with a number of options for acquiring commercial manufacturing capacity. A deterministic analysis showed that building manufacturing capacity was the highest ranking option for the scenario considered. Stochastic analyses demonstrated that this option was also the highest ranking overall.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Biopharmaceutical manufacture; Strategy selection; Decision-support; Multi-criteria decision-making; Uncertainty; Process economics

1. Introduction

Biopharmaceutical manufacture is a highly expensive and technology-intensive activity. Biopharmaceutical products are among the most technically complex to manufacture of any industry (Ransohoff, 2004). Biopharmaceuticals also take longer to manufacture than chemical entities and require production operations that are more difficult to control (Mcgurk, 2004). In recent years, it has been reported that the biopharmaceutical industry has been faced with potential shortages in manufacturing capacity with decreasing revenue potential. This is exacerbated by higher pressures from regulatory bodies (Molowa, Shenouda, & Meyers, 2001; Pisano, 1997). Although anticipated shortages may not have been realised (Ginsberg, Bhatia, & McMinn, 2002) the importance of this issue to the biopharmaceutical industry was clearly displayed in the general level of concern generated at the time. Consequently and critically, this demonstrates a link between a firm’s manufacturing operations and its long-term business strategy.

Several studies have indicated a strong link between production and business strategies (Brown, 1998; Demeter, 2003). For the biopharmaceutical industry in particular, aligning decisions with a business strategy forms a major factor in the decision of whether to make products in-house or to outsource their manufacture to a contractor (Ransohoff, 2004; Seymour & Gallibier, 2002). Because of the significant costs and resources required, the manufacturing strategy undertaken is influenced by the approach to portfolio management and drug development issues. The size of the portfolio, success rates through clinical trials, doses and expected market demand for each drug dictate the potential production requirements. The budget, cost of drug development, and speed to market required constrain the amount of capital that can be allocated to manufacturing within a specific time-period and how long the company has to acquire manufacturing capacity. All of these factors increase the difficulty facing biopharmaceutical firms, especially smaller ones, in achieving success in the industry. Given the multitude, significance, and complexity of decisions to be made for acquiring biopharmaceutical manufacturing capacity there are a surprisingly small number of published studies in the area.

There are a number of options to consider when acquiring commercial manufacturing capacity. These would be to build...
Nomenclature

**Greek letters**
- $\alpha$: total assets
- $\alpha_C$: current assets
- $\lambda$: total liabilities
- $\lambda_C$: current liabilities
- $v$: COGS at peak demand

**Subscripts**
- $i$: $i$th criterion
- $j$: $j$th alternative option
- $n$: drug $n$
- $t$: year $t$

**Symbols**
- $\alpha$: discount factor
- $h$: total sales
- $c$: coupon rate
- $C$: criterion
- $e$: attribute value
- $E$: total annual expense
- $E_C$: capital investment
- $E_{GA}$: general and administrative
- $E_R$: royalties
- $E_T$: total plant expenses to be paid
- $E_{PF}$: profits owed to partner
- $E_{F2}$: fixed capital investment
- $g$: proportion of sales to be spent on general and administrative costs
- $I$: total annual income
- $I_P$: dose per patient per year
- $I_T$: target population size
- $I_M$: market capture
- $I_A$: annual product demand
- $I_S$: estimated sales price
- $I_O$: other income
- $L$: par value of issued convertible debt
- $m$: proportion of profit payable as tax
- $M$: peak market capture
- $N$: total number of drugs
- $p$: proportion of profits owed to partner
- $p_e$: proportion of expenses that the company is obligated to pay
- $P$: profit after interest and tax
- $P_{ENPV}$: expected net present value
- $P_{NPV}$: net present value
- $P_{PV}$: annual present value
- $P_1$: profit before interest and tax
- $P_2$: depreciation
- $P_3$: taxable profit
- $P_4$: tax
- $q$: number of manufacturing personnel required
- $r$: proportion of sales that will used to pay royalty fees
- $t_D$: decay period after peak market capture
- $t_R$: ramp period from commercialisation to peak market capture
- $w$: normalised weight value
- $x$: normalised attribute value
- $Y$: project duration in years
- $z$: weight value

- $\pi_C$: cumulative transition probability
- $\pi$: transition probability
- $X_{FL}$: $k$th qualitative component of the flexibility category
- $X_{LA}$: $k$th qualitative component of the location category
- $X_{MA}$: $k$th qualitative component of the manufacturing knowledge category
- $X_{PA}$: $k$th qualitative component of the productivity category

One’s own capacity, partner with a company that can provide the required capacity, or outsource the production to a contract manufacturing organisation (CMO). Each of these options has its merits and trade-offs that need to be considered by the decision maker. Opting to build full manufacturing capacity gives the highest level of control over managerial issues and intellectual property acquisition but it also requires the largest outlay of capital and carries the highest penalty in the event of failure. Additionally, this option must be considered quite early as commercial-scale biopharmaceutical manufacturing facilities can take at least 3 years to build (Ginsberg et al., 2002). In light of this, the remaining options can be particularly attractive. Partnering with a company capable of providing manufacturing capacity allows expenses, operational and financial risk, and knowledge to be shared (George, Zahra, Wheatly, & Khan, 2001). On the downside, profits must also be shared and such alliances can be difficult to manage (Koza & Lewin, 2000). Outsourcing manufacturing responsibilities to a CMO reduces the upfront costs associated with building a facility compliant with current good manufacturing practice regulations (cGMP) (Nährí & Nordström, 2005). The cost per gram of manufacturing a biopharmaceutical with a CMO will be more expensive than building a facility and control over manufacturing activities will be compromised to some extent (Rajapakse, Titchener-Hooker, & Farid, 2005). Typically, a company will have multiple products in their pipeline and here there will be the need to consider more than one commercial manufacturing option for each product so as to arrive at the most preferable solution overall.

There are numerous contributions in literature that address the problem of expanding capacity in the chemical process industries. Internal planning decisions of a company and the recognition of market opportunities have been considered by Gupta and Maranas (2004). Cheng, Subrahmanian, and Westerberg (2005) provide an approach to capacity planning and inventory control. Oh and Karimi (2004) approach capacity expansion through a focus on regulatory factors. A mathematical programming approach to capacity planning for the pharmaceutical industry was presented by Gatica, Papageorgiou,

Literature on decision-making specific to biopharmaceutical manufacture has typically focused on strategic process decisions faced by companies that are planning to build pilot-scale or commercial-scale manufacturing facilities (Farid, Novais, Karri, Washbrook, & Titchener-Hooker, 2000; Farid, S., Washbrook, & Titchener-Hooker, 2005; Lim, Washbrook, Titchener-Hooker, & Farid, 2005; Lim, Zhou, et al., 2005; Mustafa et al., 2004). In reality, there are a significant number of biopharmaceutical companies that manufacture at least some of their drugs externally (Langer, 2004; Rogers, Maranas, & Ding, 2005) and an extension of such frameworks is required to capture this aspect of the business. Such work was investigated by Rajapakse et al. (2005) who modelled the decision to build manufacturing capacity or to outsource this activity. Additionally, these tools tend to use one or two criteria to capture the value of making a particular decision and this by definition excludes other useful evaluative criteria. The usefulness of extending the criteria considered was demonstrated by Farid, S. S., Washbrook, and Titchener-Hooker (2005) who used multiple quantitative and qualitative criteria to assess the use of disposable components when building a biopharmaceutical manufacturing facility. This was supported by a framework that used multi-criteria decision-making (MCDM).

In problems involving capacity expansion there may be many conflicting criteria to be considered for decision-making. Generally, the available types of methods for reconciling between multiple objectives include: transforming the multi-objective problem into a single objective problem, the lexicographic approach, and the Pareto approach. The field of MCDM offers numerous techniques that transform multi-objective problems into single-objective ones. This often involves assigning a weight to each objective so that relatively more important objectives contribute more to the output score of each alternative. In cases where a score is defined and in its simplest application, the most preferred candidate solution will yield the optimal score. The lexicographic approach involves assignment of a preference order to each objective and then optimizing each objective in order of preference until a non-dominated solution is found (e.g., Sawik, 1997; Volgenant, 2002). The Pareto approach (e.g., Cheng, Subrahmanian, & Westerberg, 2003) compares each solution against each criterion and results in a set of Pareto optimal solutions that are said to be non-dominated. A solution is considered Pareto optimal if there is no other existing solution that can improve the value of one criterion without degradation in any other criteria. A key advantage of using lexicographic and Pareto approaches over transforming multiple objectives into a single objective is that they compare solutions according to each criterion. However, for numerous criteria it can be difficult for the decision maker to prioritise each individual objective, as in the lexicographic approach, without having a reasonable understanding of each one. Additionally, having a set of Pareto optimal solutions means that the decision maker still has the task of selecting a solution from this set. This can be especially difficult if there are more than two criteria as multiple trade-offs may exist between individual objectives. For these reasons the simpler technique of transforming multiple objectives into a single objective has been chosen for this paper.

The use of multiple criteria as a basis for analysing decisions can be supported effectively through MCDM which is one of the most well-known branches of decision-making (Triantaphyllou, 2000). Its potential ability to support the selection of strategies in biopharmaceutical manufacturing lies in the flexibility it provides to reconcile both the financial and operational concerns of the decision maker. There are many methods available in MCDM which have been used to process financial and non-financial data. For example, Platts, Probert, and Canez (2002) used the weighted sum method (WSM) to analyse the decision to invest in internal manufacturing capabilities or to outsource these activities. Additionally, Steuer and Na (2003) have published a review of 265 publications that focus on utilising MCDM to aid decision-making in financial contexts.

This paper investigates the design of a decision-support framework for the stochastic analysis of options for acquiring commercial-scale manufacturing capacity using multiple criteria and MCDM. The remainder of this paper will discuss the architecture of the decision-support framework, demonstrate this framework by application to a case study, and draw conclusions from the results obtained.

2. Decision-support framework

The decision-support framework was designed to model the financial and operational perspectives of strategies for acquiring commercial manufacturing capacity. Microsoft Excel was used for its implementation and an overview is presented in Fig. 1. A hierarchical approach was applied to facilitate a clear and detailed representation of the business and manufacturing processes involved. This approach also offers the ability to rapidly change assumptions so that a wider variety of scenarios can be captured and analysed. The framework has four elements: a biomanufacturing process model, a profit and loss model, an MCDM technique, and a set of criteria used to distinguish between the strategic options.

A decision-making scenario is comprised of a set of alternative options with each one representing a capacity acquisition strategy. Each option is defined as a collection of probability distributions to be inputted to the model. These inputs can be split into technical, commercial, business, and qualitative categories. Technical inputs describing the manufacturing capabilities, such as the fermentation titre and the overall process yield, are directed to a biomanufacturing cost model. These technical inputs are used to generate data such as the cost of goods sold per gram (COGS) and the capital investment required. Commercial inputs include the market capture and the market lifetime of each drug. These inputs are used to define the production
Fig. 1. An overview of the MCDM framework.

Fig. 2. The biomanufacturing process model. Abbreviations: DSP, downstream processing; QC/QA, quality control/quality assurance; PCI, fixed capital investment; COGS, cost of goods sold. Demand is the annual product demand determined from the market analysis and depends on market size, market capture and the drug dosage per patient per year. Titre refers to the titre of crude product (g/L) that is expected to be achieved in the fermenter. The DSP yield is the overall yield after all downstream processing steps have been completed. The batch success rate refers to the likelihood of batch success given the chances of contamination or equipment failure. The manufacturing operations and the ancillary tasks have been modelled to determine estimates of utilisation of major cost components. Utilisation estimates are combined with an extensive cost database to determine the FCI and COGS values.
capacity required and expected sales revenue. Inputs from the business define the structure of any contractual agreements such as royalty payments and the terms of any partnerships formed. Qualitative aspects of each option include the suitability of the location, control and flexibility over the projects considered, and the expected acquisition of manufacturing knowledge. Qualitative scoring is used to describe most of the operational aspects. The outputs of the biomanufacturing cost model are directed to the profit and loss model. The outputs of the profit and loss model determine the values of the financial criteria used to discriminate between the alternative options. Subjective operational inputs are fed directly to their corresponding criteria. Details of the biomanufacturing model and the profit and loss model will be discussed in further detail in Sections 2.1 and 2.2, respectively.

2.1. Biomanufacturing process model

The framework used to model the manufacture of biopharmaceutical products (Fig. 2) is based on work reported by Farid et al. (2000), Farid, S., et al., 2005 and Lim, Washbrook, et al. (2005), Lim, Zhou, et al. (2005). Fig. 2 represents a typical manufacturing route for monoclonal antibodies. Included in the model are the main process and ancillary tasks involved in the manufacturing process as well as equations for calculating the utilisation of equipment, materials, utilities and labour. The cost calculations are supported by a database of unit costs for equipment and materials used in the manufacturing process.

The overall input parameters to the model are the annual demand, the expected fermentation titre, the overall product yield and the anticipated success rate of each batch. Each unit operation has a process model comprising of design equations and mass balances. These are used to size equipment, determine the composition of the output streams and the amount of materials required (e.g. chromatography buffers). More details on these models can be found in Farid et al. (in press). Equipment sizes are determined by matching processing requirements such as volume to a database of equipment dimensions available at the time of writing. If the processing requirements require an equipment size that is in excess of available equipment dimensions then the sizes of equipment allocated will be such that all units will be of equal size with their number clearly being sufficient to handle process requirements. The outputs from this model are the fixed capital investment for the manufacturing plant as well as the COGS. The fixed capital investment was estimated by the Lang Factor method (Lang, 1948) which correlates this figure to the equipment cost. The COGS model employed is shown in Table 1.

2.2. Profit and loss model

A profit and loss model was built to process income and expense data (Table 2) and to generate values for the decision-making criteria. In addition to the business revenues achieved from the sale of therapeutics other sources of finance are also modelled. These additional sources of finance include an initial public offering (IPO) and the issuance of convertible debt. It is assumed that this will be in the form of convertible bonds which are issued in the same year as the IPO. The holder of any convertible bonds can convert them to common stock of the issuing firm according to conditions specified by the firm prior to issuing the bonds (Kimura & Shinohara, 2006). The static conditions are the term of the bond, its coupon, the probability of conversion and the conversion premium which is the amount the bond holder must pay the firm in order to convert. The only variable condition is the year of conversion, providing this happens, which is modelled as a triangular distribution. Additionally, the year of conversion is expected to coincide with the successful commercialisation of the first drug which is reflected in the probability distribution. The amount of finance required to be raised through the issue of this instrument is calculated. The model assumes that there are no possibilities of negative cashflows. If the cashflow in any year for any scenario falls below $5 million then enough is raised to correct this with an upper limit of $10 million for the same year. It is assumed that there are no restrictions to issuing convertible debt at the time of need.
The criteria used in the framework (Table 3) were chosen to represent a broad spectrum of values that may be important when considering detailed financial and operational perspectives. Also, each criterion is intended to capture a certain aspect of each option over its lifetime from an overall perspective. In the financial category five categories of criteria are considered: profitability, asset utilisation, liquidity, long-term solvency, and capital structure. In the operational category four categories were considered: productivity, the suitability of the location, control and flexibility over the projects considered, and the expected acquisition of manufacturing knowledge. The qualitative component of the productivity category includes the ease of training manufacturing personnel. The location category includes the ability of the local infrastructure to support operations and logistics, and the access to qualified manufacturing personnel. The flexibility category includes the ability to control its drug manufacture projects, the ease of expanding operations, and the ease of consolidating manufacturing operations. The manufacturing knowledge category includes the readily available manufacturing expertise, the number of company personnel assigned to manufacturing, the potential to acquire manufacturing knowledge, and the control over any manufacturing knowledge acquired.

When applying these criteria to a decision-making scenario a higher criterion value represents a stronger position. Some of these criteria, such as those based on costs had to be inverted to reflect stronger positions as higher values rather than as lower values.

2.4. The weighted sum method

The weighted sum method (WSM) was chosen to evaluate the criteria shown in Table 3. These criteria are not all measured in the same dimensions and thus result in a multi-dimensional decision-making scenario. For comparisons between options to be drawn meaningfully and usefully the criteria values must be converted into an equivalent set of dimensionless numbers (e.g. Triantaphyllou, 2000). Normalisation can convert the criteria values as described and a normalised criterion value can be represented as:

$$x_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}}$$

where $e_{ij}$ is the value of the $i$th criterion of the $j$th alternative option and $x_{ij}$ is the normalised rating of attribute $i$ for the alternative option $j$.

Each criterion is given a weighting that can be configured by the decision maker to reflect how important each criterion is in the decision-making process. These weight values must also be converted to an equivalent set of dimensionless numbers. The normalised weight value, $w_i$, is:

$$w_i = \frac{z_i}{\sum_{i=1}^{n} z_i}$$

where $z_i$ is the weight value of the $i$th criterion.

The score, $S_j$, generated by the WSM method can be represented as:

$$S_j = \sum_{i=1}^{n} w_i x_{ij}$$

The preferred alternative option has the highest score.
Table 3
Criteria used in the MCDM framework

<table>
<thead>
<tr>
<th>Group</th>
<th>Category</th>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial</td>
<td>Profitability</td>
<td>ENPV</td>
<td>( C_1 = \sum_{n=1}^{N} \left( \sum_{i=1}^{T} \pi_{t,n} \right) ) ( \sum_{n=1}^{T} ) ( a_{t,n} P_{t,n} ) ( (37) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total profits to total sales</td>
<td>( C_2 = \sum_{n=1}^{N} \sum_{i=1}^{T} P_{t,n} \left[ \sum_{n=1}^{T} b_{t,n} \right]^{-1} ) ( (38) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total profits to total assets</td>
<td>( C_3 = \sum_{n=1}^{N} \sum_{i=1}^{T} P_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} \right]^{-1} ) ( (39) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total profits to total equity</td>
<td>( C_4 = \sum_{n=1}^{N} \sum_{i=1}^{T} P_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} - \sum_{n=1}^{T} \lambda_{t,n} \right]^{-1} ) ( (40) )</td>
</tr>
<tr>
<td>Asset utilisation</td>
<td></td>
<td>Total sales to total fixed assets</td>
<td>( C_5 = \sum_{n=1}^{N} \sum_{i=1}^{T} b_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} \right]^{-1} ) ( (41) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total profit to total current assets</td>
<td>( C_6 = \sum_{n=1}^{N} \sum_{i=1}^{T} P_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} \right]^{-1} ) ( (42) )</td>
</tr>
<tr>
<td>Liquidity</td>
<td></td>
<td>Total current assets to total current liabilities</td>
<td>( C_7 = \sum_{n=1}^{N} \sum_{i=1}^{T} \left[ \sum_{n=1}^{T} \alpha_{t,n} \right]^{-1} ) ( (43) )</td>
</tr>
<tr>
<td>Long-term solvency</td>
<td></td>
<td>Total equity to total liabilities</td>
<td>( C_8 = \left[ \sum_{n=1}^{N} \sum_{i=1}^{T} \alpha_{t,n} - \sum_{n=1}^{T} \lambda_{t,n} \right] \left[ \sum_{n=1}^{T} \lambda_{t,n} \right]^{-1} ) ( (44) )</td>
</tr>
<tr>
<td>Capital structure</td>
<td></td>
<td>Total assets to total equity</td>
<td>( C_9 = \sum_{n=1}^{N} \sum_{i=1}^{T} \alpha_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} - \sum_{n=1}^{T} \lambda_{t,n} \right]^{-1} ) ( (45) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total sales to total equity</td>
<td>( C_{10} = \sum_{n=1}^{N} \sum_{i=1}^{T} \alpha_{t,n} \left[ \sum_{n=1}^{T} \alpha_{t,n} - \sum_{n=1}^{T} \lambda_{t,n} \right]^{-1} ) ( (46) )</td>
</tr>
<tr>
<td>Operational</td>
<td>Productivity</td>
<td>Average inverted COGS</td>
<td>( C_{11} = N \left[ \sum_{n=1}^{N} \right]^{-1} ) ( (47) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total profit to total manufacturing personnel</td>
<td>( C_{12} = \sum_{n=1}^{N} \sum_{i=1}^{T} P_{t,n} \left[ \sum_{n=1}^{T} q_{t,n} \right]^{-1} ) ( (48) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qualitative productivity score</td>
<td>( C_{13} = \sum_{n=1}^{N} \sum_{i=1}^{T} \chi_{F_{t},n} ) ( (49) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qualitative location score</td>
<td>( C_{14} = \sum_{n=1}^{N} \sum_{i=1}^{T} \chi_{L_{t},n} ) ( (50) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qualitative flexible score</td>
<td>( C_{15} = \sum_{n=1}^{N} \sum_{i=1}^{T} \chi_{F_{t},n} ) ( (51) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qualitative manufacturing knowledge score</td>
<td>( C_{16} = \sum_{n=1}^{N} \sum_{i=1}^{T} \chi_{M_{t},n} ) ( (52) )</td>
</tr>
</tbody>
</table>

Note: For symbol definitions refer to the Nomenclature section.
2.5. Uncertainty

To determine the stochastic ranking configuration, input values were specified as triangular distributions characterized by minimum, maximum and expected values. These inputs were subjected to a Monte Carlo simulation and the options were ranked according to their mean WSM score. Subsequent analyses were conducted to identify any trade-offs linked to these mean scores.

3. Case to be studied

A hypothetical case study was formulated to illustrate and examine the use of the framework in capturing the financial and operational perspectives of decision-making scenarios in the acquisition of commercial-scale biomanufacturing capacity. The example is based on a biopharmaceutical company that needs to acquire commercial manufacturing capacity and has three monoclonal antibody drug candidates in its product pipeline. One has just recently entered late stage clinical trials, whereas the others are still in the early stages of clinical testing. The first, second and third drugs are required in different levels of demand. Given the progress of the drug portfolio toward commercialisation the necessity to acquire commercial manufacturing space is urgent. The company’s initial public offering (IPO) has raised $100 million. The following possible options have been identified:

- 'Partner' option: Partner with a large pharmaceutical company and split the expenses and profits.
- 'CMO' option: Outsource all manufacturing requirements to a contract manufacturing organisation (CMO).
- 'Build' option: Build a new facility and undertake all of the manufacturing by in-house.
- 'Partner/Build' option: A hybrid option involving following through with the partnership but only for the first drug. The remaining drugs will be manufactured by the construction of a new facility.
- 'CMO/Build' option: A hybrid option involving a contract manufacturer but only for the first drug. The remaining drugs will be manufactured by the construction of a new facility.

3.1. Method

The commercial and development characteristics of the drug candidates are outlined in Table 4. The input values for the 'Part-

Table 4: Market and development characteristics of drugs in the case study

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Drug 1</th>
<th>Drug 2</th>
<th>Drug 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose (grams per patient per year)</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Annual market size (patients)</td>
<td>100,000</td>
<td>200,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Clinical trial stage</td>
<td>Phase III</td>
<td>Phase I</td>
<td>Phase I</td>
</tr>
<tr>
<td>Expected time to market entry (years)</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Expected market lifetime (years)</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Note: For each input the first, second and third numbers represent the worst case, base case and best case scenarios, respectively. Where only one number is present, that value remains constant through all scenarios.
nner', 'CMO' and 'Build' options are detailed in Table 5. These values were determined based on assumed differences in capabilities between the options. For example, the CMO option was assumed to offer the highest titres due to their specialised expertise in the manufacture of biopharmaceuticals. Advice from industry experts was solicited to ensure sensible assumptions were made. For the 'Partner/Build' option, manufacture of the first drug is modelled on the input values for the 'Partner' option and manufacture of the remaining drugs are modelled on the 'Build' option. Similarly, for the 'CMO/Build' option, manufacture of the first drug is modelled on the input values for the 'CMO' option and manufacture of the remaining drugs are modelled on the 'Build' option. The convertible debt instrument is the same across all scenarios and the input values can be found in Table 6. The expected values were used to generate a deterministic ranking configuration. By default the financial criteria and operational criteria were assigned equal aggregate weights. The weighting of each group of financial criteria is equal. Likewise, the weighting of each group of operational criteria is equal.

The base case input values were processed through the decision-support framework to generate deterministic scores. A sensitivity analysis, based on the best and worst case values, was used to identify the most influential factors. A stochastic analysis was conducted using triangular distributions based on worst, base case, and best values of manufacturing, commercial, and operational variables. Here, 1000 iterations of a Monte Carlo simulation were found to be sufficient to represent the stochastic behaviour of each option considered for the purposes of comparability.

4. Results and discussion

4.1. Base case

The deterministic results are presented in Table 7, which for each option includes the WSM score, the assigned rank, and the ENPV generated.

The Build, Partner and Partner/Build options scored higher than the average WSM score. The Build option scores highest indicating that is the most preferred option. This option generates the greatest ENPV value as well as the highest total value of assets which contribute highly in the capital structure category. Additionally, the highest degree of flexibility and the greatest expected gain in manufacturing knowledge are associated with this option.

The Partner option ranks second place. The sharing of expenses results in its total liabilities being the lowest of all the options and contributes to a high score in the long-term solvency category. The Partner option also involves sharing profits making it the least profitable option. The smaller profits result in a smaller equity value when compared with other options contributing to a low score in the capital structure category. This option also scored low in the flexibility category because the partnering company has been granted some control over the manufacturing projects.

The Partner/Build option scores lower than the Partner and Build options even though it is a combination of both. The benefits that the Partner/Build option gains from the Build option are counteracted because these benefits happen halfway into the lifetime of the project. This is relevant because the discount factor and the transition probabilities used to calculate the ENPV cashflows have the resultant effect of giving events occurring soonest the highest degree of impact on this figure. This will consequently impact the WSM score. Similarly, the fixed assets that are gained from building manufacturing capacity are used to generate highly discounted profits and hence lower equity. This results in a low score in the asset utilisation category and the capital structure category. Additionally, the Partner/Build option scores lowest in the productivity category as the labour force is not used as efficiently as with the Partner or Build options.

The CMO and CMO/Build options score the lowest overall. Contributors to these scores include the higher COGS and royalty charges associated with both of these options. This results in significantly higher current liabilities and lower equity values than any of the other options and gives rise to low scores in the liquidity and long-term solvency categories. Contracting out manufacturing obligations means that these options compromise some of the potential to build manufacturing knowledge so these options achieve lower scores in manufacturing knowledge. The CMO/Build option scores low in the productivity category as the labour force is not used as efficiently as with either the CMO or the Build options.

4.2. Sensitivity analysis

This analysis displays the maximal sensitivities across all options. For individual options these values will change as may the order of these factors. The analysis revealed that the most significant factor affecting the deterministic WSM score was the market capture achieved by the company (Fig. 3). It was found that the most critical drivers were associated with commercial factors. These are similar to results from research conducted by Rajapakse et al. (2005) and Stonebraker (2002).
4.3. Expected net present value (ENPV)

Use of the WSM proved to be highly useful in its ability to aggregate many aspects of the financial analysis and the operational evaluation. Furthermore, analysing the options under uncertainty via a Monte Carlo analysis also proved to be highly informative in revealing the expected performance of each option. Fig. 4 demonstrates that the CMO/Build option is the most profitable and has a marginally greater upside potential in generating profit. The reason why the Build option does not generate the most profit in all circumstances is mainly due to its less developed technical manufacturing capabilities. In this case study, the option to build has significantly less efficient technical manufacturing capabilities than with outsourcing to a contractor. This can create a greater barrier to generating profits in the Build option than the COGS premium and royalties charged will do in the CMO/Build option. When considering the Build and CMO options alone it can be seen that the Build option demonstrates a better performance in ENPV which is consistent with research by Ginsberg et al. (2002). Additionally, it is shown that in this scenario increasing mean ENPV is coupled with greater uncertainty of this value being achieved.

4.4. Stochastic WSM scores

The stochastic WSM score highlights the holistic value of the option to the decision maker as defined by the decision-support methodology. Fig. 5 reveals the difference in this case between an analysis-based entirely on ENPV and one based on multiple criteria. In particular, the CMO/Build option which came first in the ENPV analysis is one of the lowest scoring options in the WSM analysis. Another notable difference is that none of the ranking positions achieved with the ENPV analysis are the same with the WSM analysis. In contrast, the stochastic WSM ranking positions are the same as the deterministic WSM ranking positions (Tables 7 and 8).

Fig. 5 also highlights a distinction between the top three options, Build, Partner, and Partner/Build and the remaining alternatives. By regarding the peaks of each distribution alone, the Build option can be considered to be most preferential and more certain, but not entirely distinct from the Partner option. The Partner/Build ranks closely behind the Partner and Build options but not close enough to be preferred over either. To draw further distinction between the Partner and

<table>
<thead>
<tr>
<th>Table 8</th>
<th>Deterministic rankings of the options considered within each decision-making category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Category</td>
</tr>
<tr>
<td>Financial</td>
<td>Profitability</td>
</tr>
<tr>
<td></td>
<td>Asset utilisation</td>
</tr>
<tr>
<td></td>
<td>Liquidity</td>
</tr>
<tr>
<td></td>
<td>Long-term solvency</td>
</tr>
<tr>
<td></td>
<td>Capital structure</td>
</tr>
<tr>
<td>Operational</td>
<td>Productivity</td>
</tr>
<tr>
<td></td>
<td>Location</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
</tr>
<tr>
<td></td>
<td>Manufacturing knowledge</td>
</tr>
</tbody>
</table>
4.4.1. Risks and rewards

It is useful to quantify the impact of uncertain variables for each option on the WSM score. Here, the risk refers to the semi-standard deviation and the reward refers to the stochastic WSM score. Further, deviations above the mean value are not considered to be undesirable risk so the standard deviation which includes values both above and below the mean value, is not an appropriate metric. Fig. 6 plots the semi-standard deviation versus the reward of each option and also displays the initial expenditure for each option. It is ideal to have options in the lower rightmost quadrant. The Partner/Build option is shown to perform within this quadrant. The Build option performs approximately at the average semi-standard deviation value across the group with an above average reward.

The Build option is also shown to require the highest initial expenditure. If the company could not afford this option then it would have to reconcile between the trade-offs seen with the other options that perform above the average reward. As demonstrated in Fig. 6, these options would be the Partner and Partner/Build options. The Partner option offers a higher reward than the Partner/Build option but is also coupled with significantly higher risk.

This risk and reward analysis allows the decision maker to draw clearer distinctions between the Build and Partner options than was possible with Fig. 4 or Fig. 5 exclusively. Fig. 6 makes clear that although the rewards are above average for the Build and Partner options, the Build option incurs a significantly lower risk. The Build option has the advantage of offering over 14% lower risk with nearly a 4% greater reward than the Partner option. It is intuitive to attach little advantage to this gain in reward, however, this is not necessarily an accurate association to make. As seen in this case, a gain of nearly $200 million in ENPV is a contributor to this 4% advantage in reward.

4.4.2. Financial and operational aggregate scores

While analysing the holistic value of the WSM score reveals the overall worth of the option to the decision maker it is important to discern the factors that have been most influential in score’s determination. This is especially significant in confirming whether or not the characteristics attributable to the highest ranking option are aligned with the business strategy. Due to the structure of the MCDM-based model it is possible to break down the WSM value of each option into scores that relate specifically to financial and to operational criteria. Fig. 7 demonstrates the balance of operational to financial scores for each option in this scenario. A high financial score suggests a highly sustainable investment and likewise a high operational score suggests a high level of practicality and manageability. Hence, the most desirable option is the one that lies most to the upper rightmost quadrant. The Build option performs fully above the group average for both scores. Comparison between the Build and Partner options demonstrate that the Build option is most valuable both financially and operationally.

Under uncertainty it is also highlighted that there is greater variance with the financial aggregate scores than with the operational scores. Because of the subjective nature of scoring for many of the operational attributes it would not be accurate to say that this would necessarily translate to a lower operational risk.

4.4.3. Operational to financial ratio

The relative weighting of the operational aggregate score to financial aggregate score, R, is the final analysis considered in this paper and is demonstrated in Fig. 8. The previous analyses were under the presumption that operational and financial characteristics were of equal value to the decision maker. In Fig. 8, values of R ranging from 0 to 2 were used to investigate the stability of the ranking configuration. Within this range, it was found R did not have an effect on the top ranking option demonstrating that this option dominates the other options when considering both the financial and operational aspects. It can be seen from
The results indicate that all options are financially and operationally viable but according to all analyses the Build option is most preferential. Although there were many conflicting attributes associated with the selection of any option, it is important to be aware that the Build option is the option most aligned with the company’s business strategy. Overall, the factors defining the decision proved to be its ability to exercise the option along with its willingness to accept the inherent trade-offs. It is important to remember that the results are specific to the assumptions made in the case study; for example, if tighter budget constraints were assumed, this could influence the ranking of the option to build capacity and might even rule it out completely as infeasible.

Comparing the results presented, an investigation by Rajapakse et al. (2005) contained similar investigations with regard to risk and ENPV analysis. The paper involved the comparison of the construction of a biopharmaceutical manufacturing facility and the utilisation of a CMO. Similar to the results shown here was that the CMO displayed a significantly lower performance in ENPV when compared with building a new facility. Contrary to the results demonstrated here, was that the CMO was found to be less risky than the option to build. It is difficult to compare results accurately as there are many differences between the case studies, assumptions and architectures of the models. For example, the drugs used in this case study are at later stages of development so the portfolio risk will be lower than a portfolio of drugs all entering early stage clinical trials as in Rajapakse et al. (2005). This reinforces the fact that the results are case study specific.

One issue still to be resolved is that of the weightings of the criteria and their respective categories. In the case study equal weighting was assigned to all criteria and their categories but this configuration will conflict with the preferences of a decision maker that is more aggressive on generating profits. The line of inquiry following this is then in determining a suitable weighting configuration between the group, their categories, and their individual criteria to best represent the preferences of the decision maker. This consideration is outside the scope of this paper.

Finally, the decision maker needs to be aware that, by definition, the normalisation of criteria values distorts or removes any meanings associated with their original magnitudes. Additionally, further distortion of the original magnitudes is exacerbated by the amalgamation of criteria values into a score and thus caution is required when interpreting the value of that score. As demonstrated earlier, the Build option outperformed the Partner option by a modest percentage in its WSM score but a significant gain in ENPV was a contributor in this performance.

5. Conclusions

The development of a decision-support framework to assist decision-making in strategies for the acquisition of biopharmaceutical manufacturing capacity has been presented. The framework provides a structured and transparent method of analysing such scenarios through the utilisation of MCDM and Monte Carlo simulation. Additionally, several financial and operational criteria were considered to provide a broad and detailed analysis from both perspectives. A hypothetical case study was formulated to demonstrate the usefulness and limitations of the framework.

The WSM proved to be highly suitable for data handling and for the analysis of results. The Monte Carlo simulation was valuable in highlighting the probability distributions and variance of base case values. Use of the model has highlighted that the employment of a single criterion in making strategic manufacturing decisions of this nature may not allow the decision maker to be aware of other important criteria. However, use of multiple criteria analysed under uncertainty provided a successful approach in identifying and confirming the best option. The analytical approach required highlights the complexity that can be involved in making decisions similar to the one analysed. Ultimately, a thorough and accurate analysis of financial and operational data is essential to make confident distinctions between feasible and attractive options.

Acknowledgements

The Advanced Centre for Biochemical Engineering (ACBE) is gratefully acknowledged. Support for the Innovative Manufacturing Research Centre for Bioprocessing (IMRC) housed in the ACBE by the Engineering and Physical Sciences Research Council (EPSRC) is also gratefully acknowledged. Financial support in the form of an Engineering Doctorate (EngD) studentship for E. George provided by BioPharm Services and the EPSRC is gratefully acknowledged.

References


Stochastic Combinatorial Optimization Approach to Biopharmaceutical Portfolio Management

Edmund D. George and Suzanne S. Farid*

The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, U.K.

Key strategic decisions in biopharmaceutical portfolio management include drug selection, activity scheduling, and third party involvement. Optimizing strategies is complicated by uncertainty, dependency relationships between decisions, and multiple objectives that may conflict. This paper presents the development of a stochastic combinatorial multiobjective optimization framework designed to address these issues. The framework simulates portfolio management strategies while harnessing Bayesian networks and evolutionary computation concurrently to characterize the probabilistic structure of superior decisions and evolve strategies to multiobjective optimality. This formulation is applied to a case study entailing a portfolio of five therapeutic antibody projects. Optimization was driven by two objectives that conflicted here: maximizing profitability and maximizing the probability of being profitable. Initial analysis of competing strategies along the Pareto optimal front indicated that strategies with clear differences in comprising decisions can compete with similar reward-risk profiles. Hence optimization yielded results that were not intuitive but instead suggested that flexibility between strategies can exist in such large-scale problems. A cluster analysis was used to identify the prevalence of broad and superior building blocks along the Pareto front. In-house development of drugs generally emerged as a preferred constituent of superior strategies which suggested a drive toward minimizing contracting fees, premiums, royalty charges, and losses in sales revenue to third parties; no budgetary constraints were imposed in this case study. It appeared that strategies for scheduling activities had the most overarching impact on performance. Strategies for portfolio structure appeared to have the greatest degree of flexibility relative to other strategic components.

1. Introduction

Biopharmaceutical drug development is expensive, lengthy, risky, and complex. The literature has seen published figures in excess of $800 million for developing a single drug,\textsuperscript{1,2} with 6–10 years as a typical development time.\textsuperscript{3–6} Developmental uncertainties complicate the development process by introducing the possibility that it may be necessary to terminate development of a drug before a return can ever be realized. Market-related uncertainties introduce the risk that a marketed product may not meet the revenue expectations of the developer. Hence, a major and ubiquitous problem confronting biopharmaceutical drug developers is how and when to best make and implement critical business decisions so that important rewards such as profitability are optimized. Yet more expensive, risky, and complex is the development of a portfolio of drug candidates. Within a company’s pipeline, valuing one drug at a time is not sufficient and drug developers must consider the entire portfolio under technological and market uncertainties and resource constraints.\textsuperscript{7} Here the developer must also make decisions to best construct the portfolio to optimize the management of resources and reward related trade-offs that each drug development project may introduce.

Decisions made on the portfolio will include deciding the number of comprising projects subject to those that are most promising given resource constraints. Also included is the scheduling of critical development activities subject to a prioritized order of development based on considerations such as the identification of the most promising projects and accounting for strategic windows of opportunity in targeted drug markets. Portfolio selection problems pertaining to drug development have been addressed in the literature and cover a variety of methodological features, technical problems, and scenarios. For example, see Subramanian et al.,\textsuperscript{8,9} Blau et al.,\textsuperscript{10,11} Rogers et al.,\textsuperscript{12} and Jain and Grossmann.\textsuperscript{13}

In addition to decisions made on portfolio structure, drug developers must also address acquisition of or access to infrastructure for capacity-related decisions in drug development and manufacturing endeavors. These will include important strategic decisions such as whether to integrate activities within the company, to outsource them to a contract researcher (CRO) or manufacturer (CMO), or to partner with a company that has complementary developmental, manufacturing, and marketing resources. For example, strategic outsourcing to a contractor has become a vital component of the research and development process\textsuperscript{14} and plays an increasingly important role in the operations of established and emerging pharmaceutical companies.\textsuperscript{15} There are contributions that address the issue of capacity planning in the pharmaceutical industry using optimization-based approaches of which the majority pertains to drug developers who have or plan to have integrated in-house development and manufacturing infrastructures. There are some who develop frameworks for wider settings which consider third party developers and manufacturers. These include Rogers and Maranas,\textsuperscript{12} Oh and Karimi,\textsuperscript{16} and Rajapakse et al.\textsuperscript{17,18}

Approaches have been reported for the development of frameworks that incorporate both the problems of portfolio management and manufacturing capacity planning simultaneously, representing an important advancement. These present more challenging and realistic large-scale optimization problems that offer vast extensions to the lines of inquiry for either of the two problems in isolation. Such approaches include those by Levis and Papageorgiou,\textsuperscript{19} Maravelias and Grossman,\textsuperscript{20} and Papageorgiou et al.\textsuperscript{21}
Figure 1. Schematic of the framework used to evaluate populations of candidate solutions and iteratively evolve superior populations of strategies.

Given the relevance and importance of contributions to large-scale portfolio and capacity optimization problems in biopharmaceutical settings, especially in modern industry, this paper addresses the development of a holistic framework for the simultaneous stochastic, combinatorial, and multiobjective optimization of biopharmaceutical research and development portfolio management alongside manufacturing capacity planning decisions. More specifically, given a set of drug candidates and their uncertain development, manufacturing, and commercial parameters, alongside an availability of external corporate bodies for development and manufacturing with their various uncertain technical characteristics, this work presents a novel method for finding the optimal structure of the drug development portfolio, the development sequence for the selected drug candidates, the schedule of critical development activities, and the itinerary of activities at specific stages that should be integrated in-house, outsourced, or partnered to maximize important multiple objectives. The overall framework is a combination of a simulation-based evaluation framework based on work by George et al.\textsuperscript{22} and a bespoke estimation of distribution algorithm\textsuperscript{23} (EDA) to iteratively evolve a population of candidate strategies. It is the aim of this paper to investigate the performance of strategies optimized via the proposed framework and to ascertain any important implications.

The organization of the remainder of this paper is now described. Section 2 details the complete model formulation for the stochastic evaluative framework and for the EDA. Section 3 presents a case specific to the production of monoclonal antibody drug candidates to exemplify the use of the framework. Section 4 describes and discusses the results pertaining to the case presented. The final section discusses any final conclusions.

2. Model Formulation

The stochastic optimization process makes use of an evaluative framework to capture the impact of decisions made by strategies in the modeled environment and to estimate their stochastic properties (Figure 1). This is coupled with an EDA that operates the optimization procedure through the machine learning of instances of decision variables that are associated with superior strategy performance. Any interdependent relationships between decision variables and their instances are data mined and are used to build a probabilistic model on which the formulation of new strategies is based. The quality of strategies is thereby iteratively improved.

The model is designed using C++ and MS Excel. The entire simulated environment is assembled in C++ as this contains the most frequently utilized calculations and the C++ language offers significant savings in computational time. MS Excel is used as the main graphical user interface as it offers a convenient means for storing, extracting, and manipulating data. The environment is interfaced with MS Excel via a dynamic link library where it is represented as a user-defined function. Decision variables are entered into the function and defined parameters are outputted. Modules for controlling the flow of data in the simulation and optimization procedures are compiled in Visual Basic for Applications (VBA), which is readily executable in the MS Excel environment.

2.1. Evaluative Framework. The framework for evaluating each candidate solution (Figure 1) models the performance of the strategy by mapping the consequences of decisions made within the simulated environment. Various facets of the drug development process and its wider commercial environment are captured within the modular structure of the evaluative framework. Each module is designed to carry out a set of related calculations, and as illustrated, these modules are networked and eventually result in the evaluation of a set of criteria. The following subsections will detail the contents of these modules.

2.1.1. Superstructure of a Candidate Solution. Consider generation $G(t)$ as a population of candidate solutions at a given iteration of the optimization procedure, $t$, within the total number
Clinical testing; C allows the framework to formulate a plan of critical activities where target identification; C to complete either by themselves or by another corporate body. Activity an integrated activity (I), outsourced activity (C), and partnered activity (H). Hence, these stages can be instantiated as the following: target identification (ID), preclinical testing (PC), phase I clinical trials (PI), phase II clinical trials (PII), phase III clinical trials (PIII), FDA review (FDA), and market approval (MKT). Thus, this generalized structure for a portfolio of five drugs.

Figure 2. Superstructure of a candidate strategy for the commercialization of a portfolio of five drugs.

of iterations, \( t_{max} \), that can potentially maximize the objectives of the decision maker. Each candidate solution, \( g \), in \( G(t) \) has a generalized structure for representing its decisions. Figure 2 demonstrates this generalized structure for a portfolio of five drugs, where the structure can be segregated into three subtypes: sequence, timing, and third party usage.

The strategy for the drug development sequence, \( D_g \), codes the drugs that are included in the portfolio and the order in which they are commercialized:

\[ D_g = \{ D_{g,1}, D_{g,2}, D_{g,3}, ..., D_{g,m} \} \quad \forall g \in G(t), \ i \in I \]  

where \( D_{g,i} \) is the drug chosen as the \( i \)th drug in the drug development sequence by the \( g \)th candidate solution, and \( I \) is the total number of drugs in the drug development portfolio. It is clear that \( i \) must be any drug not already chosen as part of \( D_g \).

The timing strategy for the \( i \)th drug in the development sequence according to the \( g \)th candidate solution, \( T_{g,i} \), commences development of drug \( i \) according to when drug \( i \) begins the timing for a particular stage of development. These stages can be instantiated as the following: target identification (ID), preclinical testing (PC), phase I clinical trials (PI), phase II clinical trials (PII), phase III clinical trials (PIII), FDA review (FDA), and market approval (MKT). Thus,

\[ T_g = \{ T_{g,1}, T_{g,2}, T_{g,3}, ..., T_{g,l-1} \} \quad \forall g \in G(t), \ i \in I \]  

\[ T_{g,i} = \{ ID, PC, PI, PII, PIII, FDA, MKT \} \quad \forall g \in G(t), \ i \in I \]  

where \( T_g \) is the set of timing strategies for all drugs in the portfolio.

The corporate strategy belonging to the \( g \)th solution for the \( i \)th drug at the \( j \)th development activity, \( C_{g,i,j} \), has three possible instances for each stage: in house or otherwise referred to as an integrated activity (I), outsourced activity (C), and partnered activity (P). Hence,

\[ C_{g,i} = \{ C_{g,i,1}, C_{g,i,2}, C_{g,i,3}, ..., C_{g,i,m} \} \quad \forall g \in G(t), \ i \in I \]  

\[ C_{g,i,j} = \{ C_{g,i,j,1}, C_{g,i,j,2}, C_{g,i,j,3}, ..., C_{g,i,j,n} \} \quad \forall g \in G(t), \ i \in I, \ j \in J \]  

where \( J \) is the number of development activities. This format allows the framework to formulate a plan of critical activities to complete either by themselves or by another corporate body. The development activities, \( C_{g,i,j} \), are structured as follows: \( C_{g,i,1} \), target identification; \( C_{g,i,2} \), preclinical testing; \( C_{g,i,3} \), phase I clinical testing; \( C_{g,i,4} \), phase II clinical testing; \( C_{g,i,5} \), phase III clinical testing; \( C_{g,i,6} \), manufacturing for phase I clinical trials; \( C_{g,i,7} \), manufacturing for phase II clinical trials; \( C_{g,i,8} \), manufacturing for phase III clinical trials; and \( C_{g,i,9} \), commercial manufacturing.

Structuring the strategies \( D_g \) and \( T_g \) is straightforward. Each instance within \( D_g \) can only refer to one drug, while any combination of timing instances for each \( T_{g,i} \) in \( T_g \) is permissible. Structuring the corporate strategy requires the application of some rules to avoid the formulation of nonsensical strategies. The first set of stipulations restricts the company from breaking and resuming contracts with partners:

If \( C_{g,i,1} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 2, 3, 4, 5, 6, 7, 8, 9 \} \)  

If \( C_{g,i,2} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 3, 4, 5, 6, 7, 8, 9 \} \)  

If \( C_{g,i,3} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 4, 5, 6, 7, 8, 9 \} \)  

If \( C_{g,i,4} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 5, 7, 8, 9 \} \)  

If \( C_{g,i,5} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 8 \} \)  

If \( C_{g,i,6} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 9 \} \)  

If \( C_{g,i,7} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 6 \} \)  

If \( C_{g,i,8} = P \) then \( C_{g,i,j} = P \) for \( j = \{ 5 \} \)  

This assumes that the partner is required for both clinical trial development and manufacturing. A second set of stipulations prevents the breaking of outsourcing contracts on clinical trial and clinical manufacturing activities:

If \( C_{g,i,3} = C \) then \( C_{g,i,j} = C \) for \( j = \{ 4, 5 \} \)  

If \( C_{g,i,4} = C \) then \( C_{g,i,j} = C \) for \( j = \{ 5 \} \)  

If \( C_{g,i,5} = C \) then \( C_{g,i,j} = C \) for \( j = \{ 7, 8 \} \)  

If \( C_{g,i,6} = C \) then \( C_{g,i,j} = C \) for \( j = \{ 8 \} \)  

It is assumed that contracting for clinical trial testing and contracting for manufacturing will be completed by separate companies.

These stipulations have an added benefit of reducing the decision space and computational time. Without them each drug would have a total decision space for \( C_{g,i,j} \) of \( 3^9 \) solutions over its nine constituent decision variables; now this is reduced to 207 solutions. To support computational efficiency, \( C_{g,i,j} \) is programmed as a single variable having 207 instances, rather than as nine \( C_{g,i,j} \) variables with each having three instances. This format means that \( C_{g,i,j} \) can be instantiated from a database of permissible combinations of each \( C_{g,i,j} \), as opposed to having an active procedure that validates and corrects the structure of \( C_{g,i,j} \) for each \( g \) in \( G(t) \) over each \( t \) in \( T \).
2.1.2. Stochastic Variables. To simulate the uncertain environment, the model accounts for the stochastic nature of a range of uncertain variables. In this paper, each stochastic variable is characterized by a triangular probability distribution based on the specification of maximum, minimum, and most likely values. The types of variables included in the model formulation are costs, commercial factors, and manufacturing capabilities. Each variable bears relevance to a single element of \( C_{k,j} \), and a separate distribution must be defined for each possible instance of the relevant element. Thus, for a portfolio of five drugs, individual distributions of 99 stochastic variables must be defined. Depending on the structure of \( g \), the relevant stochastic variables will be selected for sampling. Cost variables directly influence decisions made within \( D_k \). Commercial factor variables impact the structure of \( D_k \) and the instantiation of \( C_{k,j} \), and hence the choice of drugs as well as third party involvement for commercial manufacture. Manufacturing capability variables will influence decisions made on the instances of \( C_{k,j} \) for \( 6 \leq j \leq 9 \), and thus the involvement of third parties in clinical and commercial manufacturing.

2.1.3. Biomanufacturing Model. The framework used to model the manufacture of biopharmaceutical products is based on work reported by George et al., Farid et al., and Lim et al. Included in the model are the main process and ancillary tasks involved in the manufacturing procedure. The main input parameters to the model are the annual demand to be met by the facility, the expected fermentation titer, the overall product yield, and the probability of achieving success for a single batch fermentation. Equations for calculating the utilization of equipment, materials, utilities, and labor are also included. Cost calculations are supported by a database of unit costs for equipment and materials used in the manufacturing process.

2.1.4. Dependencies. Within a set of decisions it is possible that at least one of these decisions can have an impact on the performance of remaining decisions that are yet to be executed. In the real world, this can be a consequence of making decisions that affect the utilization of the same tangible or intangible resource. The framework recognizes three such contexts of dependency where this type of impact may occur: contractual, revenue, and manufacturing cost. Contractual dependencies refer to the utilization of a third party for manufacturing or research activities. Here, the premise is that the longer the period for which a third party is used and the greater the number of activities that it is involved in, the more favorable the rates it charges become. Additionally, for each third party there is a minimum charge that it will not breach. This clearly affects how the corporate relation strategy performs in the model.

Revenue dependencies refer to the impact that constituents of the company’s drug portfolio create when competing within the same market. When multiple drugs compete in the same market, each drug can suffer from reduced returns in comparison to what it might have achieved if commercialized in the absence of competing drugs, given the same commercial environment. It is also possible that multiple drugs can enhance the sales revenue of each other if they have complementary applications. In the model, the impact of this dependency is realized through \( D_k \) and \( T_k \). The configuration of \( T_k \) is important here because, in the absence of other dependencies affected by these strategies, it may be more beneficial to stagger the development of competing drugs so that there are periods of reduced or no competition. Also, the performance of these strategies will depend on the revenue-related penalties or benefits of having competing or complementary drugs in the marketplace.

Capital dependencies are modeled here as affecting the capital expense required for manufacturing a drug, due to the sharing of resources for structurally similar drugs. Where such drugs are being manufactured, it may be possible to use some part of the same manufacturing facilities, hence reducing the overall capital requirements. The impact of this dependency is reliant on the capital savings that can be realized within a group of drugs. This dependency is affected by the strategies for the structure of the drug portfolio, where the choice of structurally similar drugs supports reduced capital expense requirements. The strategy for the order in which drugs are commercialized is also important because drugs commercialized later in the pipeline may require less capital. Additionally, these strategies must reconcile the penalties and benefits of their impact on capital and revenue dependencies.

2.1.5. Timeline. The timeline module contains calculations for producing the entire schedule of activities for the portfolio. The inputs for the calculations are \( D_k \) and \( T_k \), and the time it takes to perform each activity in the process of drug development and commercialization. The timeline is used to instruct the profit and loss module when to account for the various elements on income and expense.

Estimation of Distribution Algorithm. Estimation of distribution algorithms (EDAs), are similar to a class of optimization algorithms known as genetic algorithms (GAs) and have been shown to be a class of promising approaches in solving combinatorial optimization problems. Like GAs, EDAs function by iteratively evolving a population of candidate solutions to the problem until a termination criterion is satisfied. A main driver by which GAs are thought to achieve this is the manipulation of building blocks. Building blocks are a central concept in GA theory, where superior building blocks are expected to be key components of superior solutions. They can be any set of instantiations of decision variables present in any set of candidate solutions. Additionally, a superior building block can be considered as a set of instances of decision variables that work together to support the performance of any set of superior candidate solutions in the objective space considered. The instances of decision variables forming a building block may also exhibit a dependency relationship where added value in superiority is achieved by the holistic presence of a set of instances rather than necessarily through individual incremental increases in performance contributed by each constituent of the building block. EDAs are designed to recognize superior building blocks explicitly by constructing probabilistic models of the states of decision variables that are present in superior solutions. Examples of simple EDAs in literature include the univariate marginal distribution algorithm (UMDA) and the compact genetic algorithm (cGA). Examples of more complex EDAs include the Bayesian optimization algorithm (BOA) and the multiobjective hierarchical Bayesian optimization algorithm (mohBOA).

In this paper, and like the BOA, the data structure used as the framework for the probabilistic model is a Bayesian network. A Bayesian network is an annotated directed graph that encodes probabilistic relationships, and their use here derives from the fact that artificial intelligence researchers have used them to encode expert knowledge. The topology of the network is learned directly from the structure of top-performing candidate solutions in \( S(t) \) and is then randomly sampled via the conditional probabilities that it encodes to generate \( G(t+1) \).
Estimation of Distribution Algorithm

Initial population:
Let the population of candidate solutions be $G(t)$.

$t = 1$
Randomly generate the initial population of candidate solutions, $G(1)$.

Let $t_{\text{max}}$ be the maximum number of populations to be evolved.

For $t = 1$ to $t_{\text{max}}$:

Evaluation:
Evaluate the stochastic properties of each $g$ in $G(t)$.

Selection:
Use the fast nondominated sorting and crowding distance algorithms to select the top 50% of solutions, $S(t)$, from $G(t)$.

Clustering of the objective space:
Use the k-means clustering algorithm, separate $S(t)$ into $k$ clusters.

For each cluster, $S(i, t)$:

Probabilistic modeling:
Construct the Bayesian network, $B(i, t)$, using a hill climbing procedure to optimize the Bayesian Dirichlet metric over $R(i, t)$.

Sampling of the probabilistic model:
Generate a new set of strings $A(i, t)$ by randomly sampling the joint probability distribution encoded by $B(i, t)$. The number of strings to be generated will be twice the original cluster size.

Population regeneration:
Generate the new population $G(t+1)$ by replacing all strings in $G(t)$ with all strings in $A(i, t)$.

Figure 3. Pseudocode for the EDA.

Table 1. Commercial Characteristics of Available Drug Candidates

<table>
<thead>
<tr>
<th>drug candidate</th>
<th>drug group</th>
<th>annual demand (kg/year)</th>
<th>CAGR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>Tr(120,250,380)</td>
<td>Tr(1.00,1.01,1.02)</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>Tr(100,200,300)</td>
<td>Tr(1.00,1.01,1.02)</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>Tr(80,150,230)</td>
<td>Tr(1.00,1.01,1.02)</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>Tr(50,100,150)</td>
<td>Tr(1.00,1.01,1.02)</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>Tr(100,200,300)</td>
<td>Tr(1.01,1.03,1.04)</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>Tr(80,150,230)</td>
<td>Tr(1.01,1.03,1.04)</td>
</tr>
<tr>
<td>G</td>
<td>2</td>
<td>Tr(50,100,150)</td>
<td>Tr(1.01,1.03,1.04)</td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>Tr(80,150,230)</td>
<td>Tr(1.02,1.05,1.06)</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>Tr(50,100,150)</td>
<td>Tr(1.02,1.05,1.06)</td>
</tr>
<tr>
<td>J</td>
<td>3</td>
<td>Tr(100,200,300)</td>
<td>Tr(1.02,1.05,1.06)</td>
</tr>
</tbody>
</table>

Table 2. Duration and Cost Information for Various Phases of the Drug Development Process

<table>
<thead>
<tr>
<th>phase of development</th>
<th>duration (years)</th>
<th>cost ($\text{MM}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>target identification</td>
<td>1</td>
<td>Tr(3.5,8)</td>
</tr>
<tr>
<td>preclinical</td>
<td>2</td>
<td>Tr(20,35,50)</td>
</tr>
<tr>
<td>phase I clinical trials</td>
<td>1</td>
<td>Tr(5,15,20)</td>
</tr>
<tr>
<td>phase II clinical trials</td>
<td>2</td>
<td>Tr(15,25,35)</td>
</tr>
<tr>
<td>phase III clinical trials</td>
<td>3</td>
<td>Tr(45,85,125)</td>
</tr>
<tr>
<td>scale-up synthesis</td>
<td>1</td>
<td>Tr(3,5,8)</td>
</tr>
<tr>
<td>formulation</td>
<td>1</td>
<td>Tr(5,10,15)</td>
</tr>
<tr>
<td>commercial preparation</td>
<td>1</td>
<td>Tr(1,2,3)</td>
</tr>
<tr>
<td>marketing</td>
<td>1</td>
<td>Tr(2,4,6)</td>
</tr>
<tr>
<td>FDA review</td>
<td>1</td>
<td>Tr(2,4,6)</td>
</tr>
<tr>
<td>market lifetime</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>ramp time to peak market: $I$</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
<tr>
<td>ramp time to peak market: $C$</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
<tr>
<td>ramp time to peak market: $P$</td>
<td>Tr(2,1,3)</td>
<td>-</td>
</tr>
<tr>
<td>decay time after market expiry: $I$</td>
<td>Tr(2,1,3)</td>
<td>-</td>
</tr>
<tr>
<td>decay time after market expiry: $P$</td>
<td>Tr(3,4,5)</td>
<td>-</td>
</tr>
</tbody>
</table>

"The cost of producing a drug for market is determined by the biomanufacturing cost model.

EDA used here (Figure 3) is largely based algorithmic concepts found in the mohBOA.

More detailed information regarding the selection of superior strategies using fast nondominated sorting and crowding distance, clustering of the objective space using k-means clustering, and issues concerning practicalities of constructing Bayesian networks can be found in the Appendix A, Appendix B, and Appendix C, respectively.

3. Case Study Description

A hypothetical case was formulated to illustrate and examine the ability of the framework to discover optimal strategies for performance against multiple objectives in an uncertain environment. In this case a biopharmaceutical company has 10 monoclonal antibody drug candidates available for development but can only choose five. It needs to know which drug candidates should be chosen, their order of their development, the timing schedule of development activities, and which corporate bodies should be assigned to each development activity. The optimization model is considerate of the commercial characteristics of drug candidates (Table 1), durations and costs associated with various stages of the drug development (Table 2), and technical manufacturing characteristics for each corporate body (Table 3). As seen in Table 1, there are three groups of indication that these candidates belong to. Annual demand figures and compound annual growth rates (CAGRs) are based on realistic figures. The dependencies for revenue and capital expense are detailed in Table 4. Specifications for contractual dependencies and technical probabilities of success are displayed in Table 5. An example of the complexity of the decision space concerning the five-drug portfolio is discussed. Overall, each strategy consists of 54 decision variables: five for the selection of drugs for the portfolio, four for timing the commencement of development activities for the next drug in the pipeline, and nine assignments of corporate bodies to critical activities for each of the five drugs. Hence, the choice of portfolio structure has $10 \times 9 \times 8 \times 7 \times 6$ possibilities that are combined with
Table 3. Technical Information for Manufacture of Drug Candidates According to Clinical Phase and Corporate Body

<table>
<thead>
<tr>
<th>corporate body</th>
<th>development phase</th>
<th>whole process yield (%)</th>
<th>fermentation titer (g/L)</th>
<th>batch success probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>phase I</td>
<td>Tr(0.25,0.35,0.50)</td>
<td>Tr(0.20,0.30,0.50)</td>
<td>Tr(0.80,0.60,1.00)</td>
</tr>
<tr>
<td>C</td>
<td>phase I</td>
<td>Tr(0.45,0.55,0.85)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>P</td>
<td>phase I</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.60,0.75,1.20)</td>
<td>Tr(0.65,0.85,1.00)</td>
</tr>
<tr>
<td>I</td>
<td>phase II</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.30,0.40,0.60)</td>
<td>Tr(0.65,0.85,1.00)</td>
</tr>
<tr>
<td>C</td>
<td>phase II</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
<tr>
<td>P</td>
<td>phase II</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>I</td>
<td>phase III</td>
<td>Tr(0.40,0.50,0.75)</td>
<td>Tr(1.50,2.00,3.00)</td>
<td>Tr(0.75,1.00,1.00)</td>
</tr>
<tr>
<td>C</td>
<td>phase III</td>
<td>Tr(0.55,0.75,1.00)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
<tr>
<td>P</td>
<td>phase III</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(0.80,1.00,1.50)</td>
<td>Tr(0.70,0.90,1.00)</td>
</tr>
<tr>
<td>C</td>
<td>market</td>
<td>Tr(0.50,0.70,1.00)</td>
<td>Tr(0.50,0.70,1.00)</td>
<td>Tr(0.75,1.00,1.00)</td>
</tr>
<tr>
<td>P</td>
<td>market</td>
<td>Tr(0.45,0.60,0.90)</td>
<td>Tr(1.20,1.50,2.30)</td>
<td>Tr(0.75,0.95,1.00)</td>
</tr>
</tbody>
</table>

Table 4. Specification of Dependencies As Related to the Number of Drugs from the Same Group within the Chosen Drug Development Portfolio

<table>
<thead>
<tr>
<th>no. of drugs</th>
<th>% of full revenue</th>
<th>% of full capital</th>
<th>% of full COGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>78</td>
<td>85</td>
</tr>
</tbody>
</table>

*COGS, cost of goods sold.

Table 5. Specification of Contractual Dependencies and Stagewise Probabilities of Success

<table>
<thead>
<tr>
<th>development stage</th>
<th>CRO</th>
<th>CMO</th>
<th>partner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>group 1</td>
<td>group 2</td>
<td>group 3</td>
</tr>
<tr>
<td>target identification</td>
<td>6</td>
<td>60</td>
<td>0.90</td>
</tr>
<tr>
<td>preclinical trials</td>
<td>8</td>
<td>10</td>
<td>1.50</td>
</tr>
<tr>
<td>phase I clinical trials</td>
<td>9</td>
<td>12</td>
<td>0.50</td>
</tr>
<tr>
<td>phase II clinical trials</td>
<td>10</td>
<td>14</td>
<td>0.70</td>
</tr>
<tr>
<td>phase III clinical trials</td>
<td>16</td>
<td>50</td>
<td>0.95</td>
</tr>
</tbody>
</table>

A decrease in royalty rate applies per additional drug and is measured in percentage points. Royalty rates decrease by 1% with the CRO, and by 2% with the CMO or partner for each additional drug development project for which they are involved. A minimum royalty rate applies of 3% with the CRO, 5% with the CMO, and 40% with the partner. Groups 1, 2, and 3 refer to drug groups 1, 2, and 3.

4. Results

The case study results are discussed in the following sections, focusing on analyzing competing strategies along the Pareto front generated and identifying trends along the frontier using cluster analysis.

4.1. Pareto Front Progression. In this section the performance of strategies in the final population is examined in terms of their ability to generate profits, satisfy multiple objectives, and provide an acceptable risk profile. The analysis highlights the challenges that exist when pursuing multiple objectives in the context considered.

The progression of the EDA in discovering the Pareto front is demonstrated in Figure 4, where the mean positive NPV (net present value) is a measure of reward and p(NPV > 0) is a measure of risk. As might be expected, random initialization of the population of strategies has resulted in a generally even dispersion of the mean positive NPV and p(NPV > 0) performance attributes over the objective space. Interestingly, what can immediately be seen is that it is possible to construct a portfolio of drugs with a significantly higher probability of attaining a profit than any of its individual comprising drugs. Here drugs with a probability of reaching market that is less than 0.25 have been pooled in such a way that offers the decision maker p(NPV > 0) value in excess of 0.70. Progression of the algorithm in the objective space appears to be considerably more constrained in the p(NPV > 0) dimension than it is in the mean positive NPV dimension. Progression of the evolutionary process has been accompanied by some degree of success at each generation. It is apparent that the strategies feasible in the objective space have generally converged toward a clearly defined nondominated frontier, the Pareto front. The Pareto front is wide enough to be inclusive of p(NPV > 0) values that might be realistically sought after by a decision maker in this setting. It is also apparent that for this problem the machine learning mechanisms have greatly reduced the variability encoded in the conditional and probabilistic models of superior strategies. It should be noted that the ability of the algorithm to do this is...
dependent on whether a new set of building blocks is able to clearly distinguish itself as superior to any existing alternatives in the evolutionary process. At some stages, success in some regions of the objective space is accompanied by deterioration in other regions. This deterioration refers specifically to the nonpopulation of certain regions that were previously populated. In this particular case, it was considered appropriate that the convergence of the algorithm was tracked visually, principally because the objective space has only two dimensions. The optimization procedure was considered to be converged when general progression of the Pareto front was insignificant, leading to an unfavorable loss of regions along the Pareto front. Some deterioration was tolerated in regions where rational decision makers were unlikely to seek strategies. An example of such a region would be where $p(NPV > 0)$ is less than 0.20 or where the mean positive NPV is less than $200 million. Further analysis indicated that progression beyond 17 generations for this case study leads to a severe deterioration of the Pareto front. The superior strategies selected from the 17th generation (Figure 5) show that a nondominated frontier exists in the approximate region $0.18 < p(NPV > 0) < 0.75$. The frontier indicates that a definite trade-off exists here between maximizing the mean positive NPV to be generated by a particular strategy and maximizing the probability of attaining a positively valued profit, $p(NPV > 0)$. At no point along the frontier can the maximization of mean positive NPV and $p(NPV > 0)$ be aligned. Hence, improving $p(NPV > 0)$ means accepting a degradation in mean positive NPV. Also, this forces the decision maker into the process of selecting strategies that begins with deciding upon minimum acceptable levels of profit and probability of profitability. The region $0.50 < p(NPV > 0) < 0.65$ offers the least trade-off along the frontier when searching for strategies that improve $p(NPV > 0)$. In practical terms, not all of the frontier will be appealing to the decision maker as there will be strategies that either offer probabilities of success or mean positive NPV values that are too low to be considered for further action.

4.2. Analysis of Competing Strategies. A close examination of competing strategies along the frontier indicates that they can have similar reward–risk characteristics while demonstrating marked differences in key decisions relating either to the portfolio structure (drug selection), timing, or third party strategies. Four strategies along the nondominated frontier have been selected to illustrate this and demonstrate how a decision maker might rationalize the choice between two strategies of similar risk and reward as visualized in Figure 5.

S1 and S2 are strategies that generate the greatest mean positive financial reward but also the greatest risk of being unprofitable. Although these competing strategies have very similar third party strategies and portfolio structures, they differ

![Figure 4. Progressive discovery of the Pareto front: (a) 1st generation, (b) 3rd generation, (c) 7th generation, (d) 10th generation, (e) 14th generation, and (f) 17th generation.](image-url)
drugs. Each strategy develops almost the same portfolio of drugs, drug. Resultantly, the time it takes to complete development of its drugs with lengthy development intervals between each profits from either of the first two drugs are used to fund the shorter intervals between the developments of its final three has a shorter development time of 26 years because of the commercial manufacturing. The first two drugs are either developed within 3 years of each other, as in SI, or simultan­uously, as in S2, and developing the remaining drugs once
neously, as in S2, and developing the remaining drugs once

| Table 6. Structure of the S1 Portfolio Development Strategy |
|---|---|---|---|---|---|---|---|
| 1 | F | I | P | P | P | P | P | P | 1 |
| 2 | I | C | C | C | I | I | C | FDA |
| 3 | I | C | C | C | I | I | C | FDA |
| 4 | G | I | I | C | C | C | C | FDA |
| 5 | H | C | I | I | C | I | C | FDA |

\* I, in-house activity; C, outsourced activity; P, partnered activity; C_{15}, target identification; C_{16}, preclinical studies; C_{17}, phase I clinical development; C_{18}, phase II clinical development; C_{19}, phase III clinical development; C_{20}, phase I manufacturing; C_{21}, phase II manufacturing; C_{22}, phase III manufacturing; C_{23}, market; ID, target identification; PC, preclinical testing, P/PPI/PIII, phase I/II/III clinical trials; FDA, FDA review; MKT, market approval.

| Table 7. Structure of the S2 Portfolio Development Strategy |
|---|---|---|---|---|---|---|---|
| 1 | F | I | I | P | P | P | P | P | 1 |
| 2 | I | C | C | C | I | I | C | MKT |
| 3 | I | C | C | C | I | I | C | MKT |
| 4 | B | I | I | C | C | C | C | MKT |
| 5 | H | C | I | I | C | I | C | MKT |

\* I, in-house activity; C, outsourced activity; P, partnered activity; C_{15}, target identification; C_{16}, preclinical studies; C_{17}, phase I clinical development; C_{18}, phase II clinical development; C_{19}, phase III clinical development; C_{20}, phase I manufacturing; C_{21}, phase II manufacturing; C_{22}, phase III manufacturing; C_{23}, market; ID, target identification; PC, preclinical testing, P/PPI/PIII, phase I/II/III clinical trials; FDA, FDA review; MKT, market approval.

greatly in their timing strategies as illustrated in Tables 6 and 7. Both strategies share the approach of partnering for the majority of activities for the first drug and then preferring to use a combination of in-house and outsourced development and manufacturing for the remaining drugs. The only exception to this is for drug 3 with S1, where partnering is used for commercial manufacturing. The first two drugs are either developed within 3 years of each other, as in S1, or simultane­ously, as in S2, and developing the remaining drugs once these drugs are close to the marketplace. This approach has the advantage of dividing the risk and impact of failure between two drug development projects, and it is clear that the expected profits from either of the first two drugs are used to fund the development of future projects. S1 staggers the development of its drugs with lengthy development intervals between each drug. Resultantly, the time it takes to complete development of the portfolio and subsequent market production is 35 years. S2 has a shorter development time of 26 years because of the shorter intervals between the developments of its final three drugs. Each strategy develops almost the same portfolio of drugs, with the difference being that for the fourth drug S1 develops drug G and S2 develops drug B. Much like the strategy with the first two drugs, S1 develops its third and fourth drugs together and S2 develops its fourth and fifth drugs within 3 years of each other. For S1 these two drugs are taken from the same group, whereas with S2 this is not the case. S1 and S2 generate similar performances in mean positive NPV, but a strong reason for selecting S2 exists. S2 has the additional advantage of being executed over a significantly shorter period.

Examining the differences between another pair of competing strategies illustrates how marked differences in third party strategies can also result in similar reward—risk characteristics. In contrast to S1 and S2, strategies S3 (Table 8) and S4 (Table 9) have similar timing strategies but very different third party strategies as well as portfolio structures. S3 utilizes a mixture of in-house, outsourced, and partnered activities, whereas S4 develops the entire portfolio in-house. The portfolio constructed by S3 consists of one low-demand drug, two medium-demand drugs, and two high-demand drugs that are sourced from all three groups. The first three drugs are developed first and simultaneously with the remaining two drugs developed with lengthy intervals of 9 years in between. For the second and third drugs contractors are only used for clinical development, clinical manufacturing is kept in-house, and commercial manufacturing is conducted with a partner. The fourth and fifth drugs use contracting more intensively for clinical development, but contractors are also used for clinical manufacturing where they are used during phase III trials. The total portfolio development time plus time taken during the marketing phase for S3 is 28 years. S4 develops drugs from all three groups, and its portfolio consists of two low-demand drugs, two medium-demand drugs, and one high-demand drug. Similar to S3, the first three drugs are developed in close succession and S4 also staggers the development of its final two drugs in a similar fashion. The total time for development and marketing completion of S4 is
30 years, which is similar to that taken for S3. Comparing their performances, S3 generates a marginally superior mean positive NPV than S4 and is executed over a slightly shorter period. Although marginal, the performance results suggest S3 is the superior strategy.

From sampling the strategies along the Pareto front, it is observable that their positioning in the objective space and even their occurrence within the optimized population of strategies is not intuitive. It is taken that this is due to the complexity that is inherent in the model of the biopharmaceutical drug development pathway and the complexity that governs the formulation of superior strategies. This is an important underscore because it highlights to the decision maker the value of accounting for the concept of building blocks, which is considerably difficult to achieve without the use of advanced computational tools. Another important observation is that strategies with clear differences in either drug selection, timing, or third party strategies can compete with similar reward versus risk profiles. Hence it is useful for the decision maker to identify a desirable region along the frontier and closely examine the different options that can yield the desired return and acceptable risk.

4.3. Cluster Analysis. In order to investigate if any discernible trends in strategy formulation exist among the population of superior strategies, the objective space was decomposed into three clusters. It was anticipated that an analysis of these clusters might add useful insights when considering if any particular regions of the Pareto front give rise to the prevalence of certain building blocks. As building blocks effectively compete with each other for selection, it would be useful to discover if and why any particular building blocks are emphasized.

4.3.1. Portfolio Structure. Table 10 displays the most probable constituents of strategy for each cluster. This includes the annual level of demand for the selected drug, the group that the drug belonged to, and the time after which development of the next drug would commence. The reader is reminded that group 1 drugs have a low annual growth rate, while groups 2 and 3 respectively have medium and high annual growth rates. Also, group 3 drugs have the greatest probability of achieving marketing approval while group 2 drugs have the least. Cluster 1 consists of strategies that perform highly in mean positive NPV, such as S1 and S2, and cluster 3 consists of strategies that have high values of $p(NPV > 0)$. Cluster 2 represents strategies that balance both of these performance metrics more evenly, such as S3 and S4. The clusters and their member strategies are derived by application of the k-means clustering algorithm. For cluster 1, the most probable portfolio consists of one low-demand drug, one medium-demand drug, and three high-demand drugs from all three drug groups. Strategies in this cluster tend to utilize the largest number of high-demand drugs, which assists in explaining the larger mean positive NPV values generated by these strategies. Also, a larger portion of selected drugs belong to drugs with high and medium probabilities of success. For cluster 3 at the opposite end of the Pareto front, the most probable portfolio is three medium-demand drugs and two high-demand drugs drawn from all three drug groups, which perhaps surprisingly highlights that cluster 3 strategies aim to meet a similar level of market demand as cluster 1 drugs. As cluster 3 strategies have significantly lower mean positive NPV values than cluster 1 strategies, this indicates that there are other decisions outside of portfolio structure which lower this profitability measure but may contribute to a greater $p(NPV > 0)$ value. With the exception of one drug, the portfolio opts for drugs with high and medium probabilities of success, thereby using more drugs of this type than the other clusters. This is intuitive because this cluster exhibits the greatest $p(NPV > 0)$ values. Overall, there appear to be no clear trends in portfolio structuring strategies. The presence of certain portfolio structuring decisions within certain clusters appears to align with the position in the objective space relative to other clusters, while there are others which do not. The remaining constituents of the portfolio development strategy will be investigated to discover if more overarching drivers of performance exist.

4.3.2. Timing Strategies. Decisions on timing are an important constituent of the portfolio development strategy as they are used to favorably organize cash flows. This is particularly important when having to consider the probability that a project will succeed, the financial impact of failed projects, and the impact of the discount factor when determining the mean positive NPV. In cluster 1, Table 10 shows that the prevailing strategy is to develop the first two drugs together coupled with long intervals of either 8 or 9 years between the development of the remaining drugs. Because of the probabilities of success of available drugs, it would be rational to expect only one successful drug within a portfolio of five projects. Also, it is observable that, when considering the growth and impact of the discount factor over time, cluster 1 strategies must rely considerably on the production of a successful drug from the first two projects. This assists in reasoning the low $p(NPV > 0)$ values seen in this cluster. Remaining drugs have comparatively little impact on cash flows because of the extent to which their respective cash flows are discounted. This also provides an explanation for the superior mean positive NPV values, as when one successful drug emerges from within the first two projects its profits are not greatly discounted. Also, and importantly, profits from this drug will mainly need to absorb the expense of its own development and that of the other concurrent project. The cost of developing the remaining drugs is less restrictive as these are heavily discounted, although this occurs with a relatively modest probability. The time for

---

**Table 9. Structure of the S4 Portfolio Development Strategy**

<table>
<thead>
<tr>
<th>i</th>
<th>D_i</th>
<th>C_{i1}</th>
<th>C_{i2}</th>
<th>C_{i3}</th>
<th>C_{i4}</th>
<th>C_{i5}</th>
<th>C_{i6}</th>
<th>C_{i7}</th>
<th>C_{i8}</th>
<th>T_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>PC</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>ID</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>FDA</td>
</tr>
<tr>
<td>4</td>
<td>J</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>FDA</td>
</tr>
<tr>
<td>5</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>-</td>
</tr>
</tbody>
</table>

*a, in-house activity; C, outsourced activity; P, partnered activity; C_{i1}, target identification; C_{i2}, preclinical studies; C_{i3}, phase I clinical development; C_{i4}, phase II clinical development; C_{i5}, phase III clinical development; C_{i6}, phase I manufacturing; C_{i7}, phase II manufacturing; C_{i8}, phase III manufacturing; C_{i9}, market. ID, target identification; PC, preclinical testing; PI/PII/PIII, phase I/II/III clinical trials; FDA, FDA review; MKT, market approval.*

**Table 10. Characteristics of Drug Selection and Timing Strategies in Each Cluster**

<table>
<thead>
<tr>
<th>cluster</th>
<th>characteristic</th>
<th>drug 1</th>
<th>drug 2</th>
<th>drug 3</th>
<th>drug 4</th>
<th>drug 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>demand</td>
<td>M</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>drug group</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>timing</td>
<td>ID</td>
<td>FDA</td>
<td>PIII</td>
<td>FDA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>demand</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>drug group</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>timing</td>
<td>PC</td>
<td>PC</td>
<td>MKT</td>
<td>FDA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>demand</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>drug group</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>timing</td>
<td>ID</td>
<td>FDA</td>
<td>PIII</td>
<td>PIII</td>
<td></td>
</tr>
</tbody>
</table>

ID, target identification; PC, preclinical testing; PI/PII/PIII, phase I/II/III clinical trials; FDA, FDA review; MKT, market approval; H, high; M, medium; L, low.
development until the end of marketing for this strategy is 34 years.

For cluster 2 it is shown that the first three drugs are most likely to be developed with short intervals of 1 year between them and that the time for development and marketing completion for this approach is 31 years. This strategy is a driver of the higher \( p(NPV > 0) \) values and the lower mean positive NPV values observed when making comparisons to cluster 1. Like cluster 1, the remaining drugs are likely to be developed with the longest intervals between them and are anticipated to have modest impacts on cash flow because of the magnitude of the discount factor at these stages in portfolio development. Considering this, such strategies are reliant on at least one successful drug emerging from the first three projects. Developing the first three drugs in close succession in cluster 2 bears a greater likelihood that at least one successful project will emerge from this group than from within the remaining two drugs. When a successful drug emerges from this group, because of a more favorable discount factor there is also a greater likelihood that it can also cover the expense of other failed projects than when developing two drugs in close succession. Having to effectively absorb the cost of an increased number of projects whose development costs are less discounted contributes to lower mean positive NPV values.

Strategies in cluster 3 exhibit a tendency to develop the first two drugs together followed by a period of at least 9 years; then later drugs are developed with medium-length intervals of 4 years in between them. The time for development and marketing completion most likely in cluster 3 is 27 years. This strategy relies on at least one successful drug being found in two distinct groups, that is, a group of two drugs developed together at the outset and a group of three drugs developed in close succession later in the timeline. It has already been mentioned that the impact of developing two drugs together while positioning the remaining projects to subject them to far more significant discounting results in relatively high mean positive NPV values with relatively low probabilities of success. It follows that some aspects of the development of the remaining three drugs contribute significantly to the high \( p(NPV > 0) \) values but low mean positive NPV values ultimately exhibited by strategies in this cluster. The shorter portfolio development time means that profits for each year are discounted by a significantly smaller extent than seen with clusters 1 and 2. Strategies in this cluster also take advantage of the high annual market growth rate and high probability of success of group 3 drugs and position them late in the pipeline of products to maximize their potential for generating revenues that overcome the magnitude of the discounting. This is important because if each drug is to contribute significantly to the revenue generated by the portfolio, then any successful drugs must support up to four failed projects, which helps to explain the smaller mean positive NPV values. Also, it is more likely to see at least one successful drug emerge from within the second group of three projects than from the first group of two. When this occurs, profits arise from these drugs which are heavily discounted and must also be enough to cover the expenses of failed projects which are less heavily discounted. This further erodes the potential magnitude for mean positive NPV. Overall, the trend with timing strategies is that portfolio development times tend to become progressively shorter toward the right-hand side of the frontier and that strategies make use of grouping projects more closely together. Interestingly, the presence of the discount factor appears to have a significant influence on strategy formulation. Hence it is anticipated that alternative settings for

---

**Table 11. Selected Probabilities of Third Party Strategies**

<table>
<thead>
<tr>
<th>Drug</th>
<th>D1</th>
<th>C1,0</th>
<th>C1,1</th>
<th>C1,2</th>
<th>C1,3</th>
<th>C1,4</th>
<th>C1,5</th>
<th>C1,6</th>
<th>C1,7</th>
<th>C1,8</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug1</td>
<td></td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Drug2</td>
<td></td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Drug3</td>
<td></td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Drug4</td>
<td></td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Drug5</td>
<td></td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.47</td>
<td>0.46</td>
<td>0.45</td>
</tr>
</tbody>
</table>

*For each Di in each cluster, the two most probable strategies have been displayed; I, in-house activity; C, outsourced activity; P, partnered activity; C, target identification; C, preclinical studies; C, phase I clinical development; C, phase II clinical development; C, phase III clinical development; C, phase I manufacturing; C, phase II manufacturing; C, phase III manufacturing; C, market.*
in-house as these offer the greatest potential for achieving the greatest mean positive NPV values if the drug passes all stages of testing and review. For the first drug an alternative strategy, albeit much less likely in this case, is to partner from the commencement of phase III clinical trials or from preclinical testing. Partnering from phase III clinical trials allows for some savings on the proportion of revenue that must be paid to the partner. Partnering from preclinical testing onward allows for the impact of failure to be shared, but a substantial share of any sales revenues must be paid to the partner. Across all clusters strategies tend to utilize partners much less intensively from drug 1 onward, where partners are used almost exclusively for commercial manufacturing and contractors are preferred. The use of contractors does not appear to be limited to a particular set or series of activities, and in most cases the use of contractors is a supplement to integrated activities. Finally, although the model shows that in-house activities are the most probable for optimized strategies, the presence of budgetary constraints is expected to impact such a decision. This will be explained in a later paper.38

5. Conclusions

The development of a stochastic multiobjective combinatorial optimization framework has been presented that addresses three key decisions simultaneously: portfolio management, scheduling of drug development and manufacturing, and the involvement of third parties for specific activities. Demonstrated within this work is the value carried by considering these critical strategic considerations within a unified framework that simulates and optimizes all such decisions across the entire product portfolio. A case study was used to illustrate the capabilities of the framework and also highlighted that the scope of decisions that a drug developer may be confronted with can be vast and complex. Due to the complexity of this problem, a principle contribution of this work is in demonstrating a formulation based on techniques from artificial intelligence, in particular evolutionary computation and machine learning, employed for an efficient search of the decision space and for effective traversing of the objective space.

It is proposed that biopharmaceutical product development strategies in the real world may be better analyzed when considering the impact of decisions holistically rather than only individually. One reason for this is the presence of dependencies between decisions that may impact economic relationships. Another is that it has been demonstrated that an effective strategy for portfolio development can result in a p(NPV > 0) value that is significantly greater than the probability of successful development for any singular drug in the portfolio. Hence, by considering a portfolio of multiple drugs it is possible to control its risk to some extent through careful strategic formulation, while this is not possible with a singular drug. Use of the model has highlighted that pursuit of mean positive NPV can conflict with pursuit of high p(NPV > 0) values, although this may change under different case study settings. The results of the case study lean toward suggesting the integration of all activities in-house. This can conflict with common perspectives in industry that accept such strategies with reluctance because of the uncertainty of the drug development process and the consequential impact of failure. The added presence of budgetary constraints and a range of sizes for the portfolio would serve as factors that can further exacerbate decisions in the real world and are also capable of significantly influencing results seen here; this is explored in a further paper.38 Furthermore, the optimal set of solutions is expected to be sensitive to the relative difference in manufacturing efficiencies assumed between in-house and external manufacturing.

Finally, the learning of Bayesian networks from superior solutions presented here has been shown to be effective and efficient in improving the population, and in discovering a dense and widespread Pareto front. Their effectiveness is presumed to be due to their ability to iteratively learn and exploit the structure of the problem as noted by contributions in artificial intelligence literature. A noteworthy insight from using the framework is that use of machine learning has potential for future development in solving portfolio development and capacity planning problems simultaneously.

Acknowledgment

The Advanced Centre for Biochemical Engineering (ACBE) is gratefully acknowledged. Support for the Innovative Manufacturing Research Centre for Bioprocessing (IMRC) housed in the ACBE by the Engineering and Physical Sciences Research Council (EPSRC) is also gratefully acknowledged. Financial support in the form of an Engineering Doctorate (EngD) studentship for E.D.G. provided by BioPharm Services and the EPSRC is gratefully acknowledged.

Appendix A. Selection of Superior Candidate Solutions

To facilitate the selection of superior candidate solutions, fast dominated sorting and crowding distance algorithms39 are used together. Both procedures are used to support the discovery of a widespread and densely populated Pareto optimal front. Fast dominated sorting ranks each strategy according to its degree of Pareto optimality in the objective space. A candidate solution is Pareto optimal if there is no other solution that exists which yields an improvement in any objective without yielding degradation in any other objective. To support the function of this sorting procedure, crowding distance is used to determine the proximity of a candidate solution to others in the objective space. The crowding distance measure for a candidate solution is the perimeter of the cuboid created by its two most proximal candidate solutions in the objective space. The premise is that candidate solutions situated in less densely populated areas of the objective space are considered to be more unique and thus can contain information not already discovered. Probabilistic models that are more inclusive of unique candidate solutions facilitate the use of a diverse base of information for the formulation of new superior candidates. Fast nondominated sorting is used first to select candidate solutions in order of rank and add them to the set of superior performing candidate solutions, S(t). If all the members sharing the same rank will not fit into S(t), then candidate solutions are chosen in order of those having the greatest crowding distance. Algorithmically steering the optimization procedure toward greater inclusion of candidate solutions along the Pareto front has its advantages here. First, the nature and severity of conflict between objectives can be better understood with wider coverage of the objective space. Second, a more efficient traversing of the objective space can be achieved through accelerated progress at individual iterations due to probabilistic models that model nondominated solutions more closely.

Appendix B. k-Means Clustering

Once S(t) has been populated, it is clustered so that the properties of superior candidate solutions can be analyzed for region-specific properties in the objective space. The k-means clustering algorithm40 was selected for this purpose because its ease of
The advantage of clustering the objective space is that a probabilistic model can be built for each cluster, thereby supporting the generation of a widespread Pareto front. This algorithm requires the number of clusters to be specified by the decision maker.

Appendix C. Constructing the Bayesian Network

Bayesian networks are used here to characterize the probabilistic nature of relationships between decision variables and their instances occurring in the set of candidate solutions of the zth cluster, K, via machine learning. This data structure is subsequently sampled for the generation of G(t + 1). As the objective space is partitioned into z clusters a Bayesian network, B, is constructed for each z. Each decision variable, δ^z, is represented by a node, δ^z, in the network B. Any relationships between each δ^z are represented by a directed edge drawn between them (Figure 6). The node from where the edge is directed is a parent node, while the node to which the edge is directed is its child node. That is, within each B, the instance of each child node is conditional upon the set of instances of its parents. It is important to note that all B's are acyclic, so a node cannot be directed to itself nor can it be part of a structure within B that eventually ends up being directed to itself (Figure 6).

In order to avoid this, the paper takes advantage of the nature of the problem for which this framework is developed and puts all δ^z into ascending order according to the time in the development schedule that each δ^z occurs. Hence, this sequence will be

δ^z = D^z , δ^z = C^z , δ^z = T^z , δ^z = D^z , δ^z = C^z , δ^z = T^z , ... , δ^z = T^z , δ^z = D^z , δ^z = C^z ,

where D^z is the decision variable representing the choice of the ith drug, C^z is the decision variable for the choice of corporate relations strategy for the ith drug, and T^z is the decision choice for the length of time to wait before development of drug i − 1 is commenced. The stipulation is imposed that no edge can be directed from any δ^z to any other δ^z, which occurs before it in this specified time order.

The procedure for constructing B, from K, is now described. A score-based approach is used with the Bayesian Dirichlet (BD) metric and the network is constructed with the intent to maximize this score. The BD metric is a measure of how closely B models the data in K. The construction procedure begins with an empty network and each possibility for adding the next directed edge in B, is scored. The next directed edge to be added will be the one that maximizes the value of the BD metric. This process continues until no improvement in score can be achieved. Additionally, no restrictions are imposed on the complexity of the network that can be built so that no important topological details are lost.

The BD metric for B given K and background information, ξ, is denoted by p(K, B|ξ) and is defined as

\[
p(K, B|ξ) = \frac{\prod_{n=1}^{N} \frac{\Gamma(\epsilon(\delta_{m,n}))}{\Gamma(\epsilon(\delta_{m,n}) + \epsilon(\delta_{n,m}))} \times \prod_{n,m=1}^{N} \frac{\Gamma(\epsilon(\delta_{m,n}, \delta_{n,m}))}{\Gamma(\epsilon(\delta_{m,n}, \delta_{n,m}))}}{\prod_{n=1}^{N} \Gamma(\epsilon(\delta_{n,m}))}
\]

where p(K, B|ξ) is the prior probability of B; δ_{m,n} is the nth node instantiated to the mth value in K; the product over δ_{n,m} runs over all instances of all parents of δ_{n,m}; the product over δ_{n,m} runs over all instances of all parents of δ_{n,m}; \epsilon(\delta_{m,n}) is the number of instances in K where the parent nodes of δ_{m,n} are instantiated to δ_{m,n}; \epsilon(\delta_{m,n}, \delta_{n,m}) is the number of instances in K that have δ_{m,n} equal to δ_{m,n} and \Pi_{δ_{m,n}} set to δ_{m,n}; and Γ(x) is the Gamma function where Γ(x) = (x − 1)! When the set \Pi_{δ_{m,n}} is empty, there is one instance of \Delta_{δ_{m,n}} that is equal to 0, and the number of instances is set to the number of instances of K. Additionally

\[
\epsilon(\delta_{m,n}) = \sum_{δ_{n,m}} \epsilon(\delta_{m,n}, \delta_{n,m})
\]

where \epsilon(\delta_{m,n}) is an instance of the parents of δ_{m,n}, that is summed over all instances of δ_{m,n}. The numbers \epsilon(\delta_{m,n}, \delta_{n,m}) and p(K, B|ξ) represent prior information about the problem that can be incorporated into the metric.

The extent to which the network being measured represents another network relevant to the problem is measured by p(B|ξ). There are a number of methods available for calculating p(B|ξ). In this paper all networks are treated equally; thus p(B|ξ) is set to 1. There are also a variety of methods for setting the numbers \epsilon(\delta_{m,n}, \delta_{n,m}) for example, see Buntine, Heckerman et al., and Yang and Chang. In particular, research by Yang and Chang has demonstrated that, for a range of integer values of \epsilon(\delta_{m,n}, \delta_{n,m}) between 1 and 10, and also for other methods considered for scoring Bayesian network topology, the BD metric with \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 ranked as one of the best for discovering the true structure of the Bayesian network. Accordingly, \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 is used here. Additionally, because the factorials in eq 20 can grow to unmanageably large numbers, especially within large sizes of K, its logarithmic equivalent is used.

Appendix D. Regenerating the Population

In generating G(t + 1) each δ^z is treated as a random variable that can be instantiated to any of its possible instances subject to probabilities encoded by B. Each decision variable possible δ^z is conditional upon each possible δ^z given K, and is computationally determined and referenced when needed. The process for generating one candidate solution for G(t + 1) requires iterative sampling of B.

The joint distribution encoded by B, is treated equally; thus p(B|ξ) is set to 1. There are also a variety of methods for setting the numbers p(\delta_{m,n}) for example, see Buntine, Heckerman et al., and Yang and Chang. In particular, research by Yang and Chang has demonstrated that, for a range of integer values of \epsilon(\delta_{m,n}, \delta_{n,m}) between 1 and 10, and also for other methods considered for scoring Bayesian network topology, the BD metric with \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 ranked as one of the best for discovering the true structure of the Bayesian network. Accordingly, \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 is used here. Additionally, because the factorials in eq 20 can grow to unmanageably large numbers, especially within large sizes of K, its logarithmic equivalent is used.

The joint distribution encoded by B, is treated equally; thus p(B|ξ) is set to 1. There are also a variety of methods for setting the numbers p(\delta_{m,n}) for example, see Buntine, Heckerman et al., and Yang and Chang. In particular, research by Yang and Chang has demonstrated that, for a range of integer values of \epsilon(\delta_{m,n}, \delta_{n,m}) between 1 and 10, and also for other methods considered for scoring Bayesian network topology, the BD metric with \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 ranked as one of the best for discovering the true structure of the Bayesian network. Accordingly, \epsilon(\delta_{m,n}, \delta_{n,m}) = 10 is used here. Additionally, because the factorials in eq 20 can grow to unmanageably large numbers, especially within large sizes of K, its logarithmic equivalent is used.
Each $K_t$ is proportionally represented in the generation of $G(t + 1)$. As $S(t)$ contains half the solutions present in $G(t)$, the number of solutions generated using each $B_t$ is $2K_t$.

**Literature Cited**


Received for review February 25, 2008
Revised manuscript received June 25, 2008
Accepted August 15, 2008

IE8003144