Noname manuscript No.
(will be inserted by the editor)

Optimal control of dosage decisions in controlled ovarian hyperstimulation

Miao He · Lei Zhao · Warren B. Powell

Received: date / Accepted: date

Abstract In the *controlled ovary hyperstimulation* (COH) cycle of the *in vitro fertilization-embryo transfer* (IVF-ET) therapy, the clinicians observe the patients' responses to gonadotropin dosages through closely monitoring their physiological states, to balance the trade-off between pregnancy rate and *ovarian hyperstimulation syndrome* (OHSS) risk. In this paper, we model the clinical practice in the COH treatment cycle as a stochastic dynamic program, to capture the dynamic decision process and to account for each individual patient's stochastic responses to gonadotropin administration. We discretize the problem into a Markov decision process and solve it using a slightly modified backward dynamic programming algorithm. We then evaluate the policies using simulation and explore the impact of patient misclassification. More specifically, we focus on patients with *polycystic ovaries syndrome* (PCOS) or potential, that is, the patients that tend to be more sensitive to gonadotropin administration.

Keywords Controlled ovarian hyperstimulation · Pregnancy rate · OHSS risk · Dosage decisions · Markov decision process

1 Introduction

Since the birth of the first "test-tube" baby, Louise Joy Brown, in 1978 in Oldham, England, *assisted reproductive technology* (ART) has been widely applied to help sterile couples to have their own children. *In vitro fertilization-embryo transfer* (IVF-ET) is the most commonly used fertility treatment, accounting for more than 99% of all ART therapies. IVF combines the oocyte and sperm in the laboratory environment, and ET transfers the embryo into the body of the mother. According to the "2005 Assisted Reproductive Technology (ART) Report" from the Division of Reproductive Health of Centers for Disease Control and Prevention (http://www.cdc.gov/ART/ART2005/), in 2005, 134,260 ART cycles were performed at 422 fertility clinics in the United States, resulting in 38,910 live births (deliveries of one or more living infants) and 52,041 infants. About 20,000 ART cycles were performed in China in 2004.

However, it is difficult to duplicate Louise's success, because the oocyte was produced in a natural menstrual cycle of her mother, without any stimulation to the ovaries. In a natural menstrual cycle, a woman usually produces

Miao He
Department of Industrial Engineering, Tsinghua University, China
E-mail: hem03@mails.tsinghua.edu.cn

Lei Zhao
Department of Industrial Engineering, Tsinghua University, China
E-mail: lzhao@tsinghua.edu.cn

Warren B. Powell
Department of Operations Research and Financial Engineering, Princeton University
E-mail: powell@princeton.edu

only *one* oocyte (in rare cases two oocytes but no more than that; some women can not produce even a single oocyte in a menstrual cycle). If the oocyte is of poor quality or can not be used at all, the chance of pregnancy is very low, which is more likely to happen among infertile couples. Given multiple oocytes, the probability that at least one oocyte is combined with a sperm increases. Therefore, the prevailing clinical practice for IVF-ET is to stimulate the ovaries with medicines to induce multiple oocytes, in order to greatly increase the pregnancy rate (see, for example, Barbieri and Hornstein 1999). The process of inducing multiple oocytes by medicines is called *controlled ovarian hyperstimulation* (COH) and the medicines administered are exogenous *gonadotropins* (or *gonadotrophins*). We depict the whole IVF-ET procedure with gonadotropin stimulation in Figure 1.

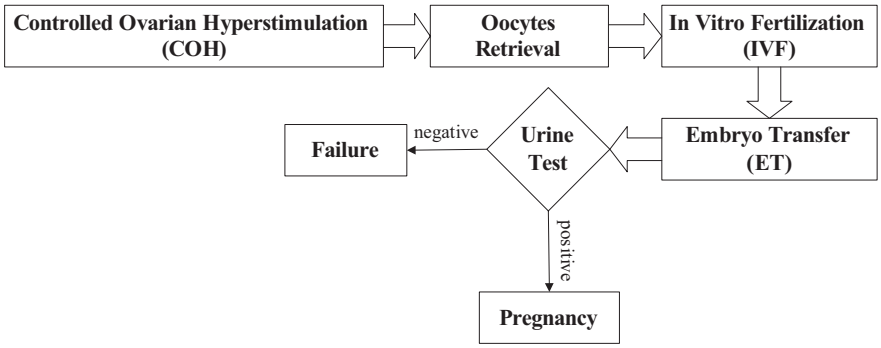


Fig. 1: A Typical IVF-ET Procedure

While the retrieval of multiple oocytes via gonadotropin stimulation in COH helps to increase the pregnancy rate, the existence of an iatrogenic disease called *ovarian hyperstimulation syndrome* (OHSS) is a constant concern in IVF-ET practices. OHSS is a serious and potentially life-threatening complication (Fauser et al. 1999). Although the true pathology is still not fully understood, it is widely accepted by the clinicians that OHSS is triggered by the administration of exogenous gonadotropins. As claimed by Delvigne and Rozenberg (2003), "Today, it is the loss of control over hyperstimulation which constitutes the OHSS."

IVF-ET therapy is highly rewarding in that it provides hope for infertile couples. Yet it is expensive and accompanied with risks. Not only does the OHSS risk jeopardize women's health, but also the treatment failures have psychological impacts on the patients. The COH cycle is an integral part of the IVF-ET therapy and affects its success. It is a stochastic, dynamic, and complicated process. Clinicians make treatment decisions (daily gonadotropin dosages, etc.) to balance the chance of pregnancy and the OHSS risk.

At the beginning of the therapy, clinicians usually (implicitly) classify patients into responsiveness/sensitivity classes (to exogenous gonadotropins) based on their medical characteristics, such as age, body mass index (BMI, defined as $body\ weight\ (kg)/stature^2\ (m^2)$), previous IVF-ET experiences, etc. In the COH cycle, clinicians need to consider each patient's individual physiological responsiveness to exogenous gonadotropin by closely monitoring her key physiological states. Clinicians base their dynamic treatment decisions (gonadotropin dosages) on guidelines from the clinical literature and (mainly) on their clinical experiences.

In the clinical literature, researchers study factors on the pregnancy rate and/or the OHSS risk using statistical analysis of clinical records. These studies provide useful guidance on clinical practices. However, most of them focus on a single medical characteristic or physiological state, yet controversial conclusions exist. Furthermore, there is a lack of research on integrating the statistical observations/conclusions into decision support tools to assist clinicians to actively respond to the stochastic and dynamic progress of the COH cycle and adjust dosage decisions to achieve more effective control over the treatment cycle and its outcomes. Such data-driven clinical decision support tools can also help to avoid subjective decisions in experience-based practices.

In this study, we propose a data-driven clinical decision support framework based on a Markov decision process (MDP) model supported by the clinical literature, expert opinions, and statistical analysis on real world clinical records, Next, we provide more details on the COH cycle and the associated OHSS risk in Section 2. We then

formulate a stochastic dynamic programming model in section 3 and describe our solution approach in section 4. In section 5, we provide our computational results. We conclude the paper in section 6.

2 Background

In this section, we provide some background on the problem in this study. In the discussion that follows, we will use "woman," "patient," and "human body" interchangeably, all of which refer to the woman undergoing the IVF-ET therapy.

Controlled Ovarian Hyperstimulation (COH)

Before the oocytes' maturity, they are called *follicles*. In a natural menstrual cycle, the endogenous gonadotropins (gonadotropins secreted by the women) surpass the threshold of a single winner follicle, but would not reach the thresholds of all others (called *recruitment*). Consequently, only the winner follicle develops into an oocyte while the growth of others is hindered (called *atresia*).

In COH, the clinicians first suppress the endogenous gonadotropins with medicines such as *gonadotropin releasing hormone analogue* (GnRH_a) or *gonadotropin releasing hormone antagonist* (GnRH_A), in order to inhibit the recruitment process, and then supplement the human body with exogenous gonadotropins to stimulate multiple follicles to grow into oocytes. The COH cycle helps the woman undergoing IVF-ET to obtain multiple oocytes to increase the chance of pregnancy. It has been an integral part of the IVF-ET therapy since the late 1970s and early 1980s.

The common exogenous gonadotropins adopted by the clinicians are *human menopausal hormone* (hMG), *follicle stimulating hormone* (FSH), or the combination of the two. Gonadotropin dosages are measured in *ampoule*. In practice, gonadotropins are administered daily.

When the diameters of the *two largest follicles* reach 18 mm, another type of gonadotropin, *human chorionic gonadotropin* (hCG), is administered to induce the final maturity of the follicles into oocytes, which marks the end of the COH cycle. The COH length varies with cycles, usually in 6-20 days (Martin et al. 2006).

Ovarian Hyperstimulation Syndrome (OHSS)

OHSS, the iatrogenic complication associated with the COH cycle, can be mild, intermediate, severe, or even life-threatening. According to Klemetti et al. (2005), the incidence of severe OHSS has been reported to vary from 0.7% to 1.7%. The incidence of severe OHSS reported by Delvigne and Rozenberg (2002) ranges from 0.5% to 5 %. OHSS can occur a few days after hCG administration (early OHSS) or later in pregnancy (late OHSS). The symptoms of severe OHSS include rapid weight gain, tense ascites, hemodynamic instability (orthostatic hypotension, tachycardia), respiratory difficulty (tachypnea), progressive oliguria, and laboratory abnormalities.

Polycystic Ovary Syndrome (PCOS)

There exists a disease called *polycystic ovary syndrome* (PCOS), characterized by 1) the presence of 12 or more follicles with 2-9 mm in diameter in a single ovary, or 2) increased ovarian volume ($> 10\ mL$), according to Norman et al. (2007). PCOS patients are *more sensitive* to gonadotropin administration than normal patients (Balasch et al. 2001, Aboulghar and Mansour 2003, Tarlatzis 2002). As a result, under gonadotropin stimulation, women with PCOS are exposed to higher OHSS risk. In the literature survey by Delvigne and Rozenberg (2002), a study reports that 63% of severe OHSS patients show ultrasonically diagnosed PCOS, while another study of 128 Belgian OHSS patients shows that 37% of them suffer from PCOS compared with 15% PCOS incidence among 256 non-OHSS patients. PCOS patients are the target group of this study.

The Tradeoff between the Pregnancy Rate and OHSS Risk

The prevailing clinical practice in IVF-ET today is to stimulate the growth of multiple follicles into oocytes to increase the chance of pregnancy, but at the same time, the clinicians need to be cautious of the OHSS risk.

On one hand, if a woman undergoing COH fails to produce multiple oocytes with satisfactory quality, the chance of pregnancy is greatly reduced and perhaps another therapy cycle is needed. Yet, the therapy is expensive. The

cost is around $7,500 per cycle in the United States (Schmittlein and Morrison 2003) and $2,500 per cycle in China (He 2004). On the other hand, if the woman suffers from (early) severe OHSS, she has to postpone the embryo transfer (ET) via *embryo cryopreservation* (to store the embryo to transfer until the woman recover from OHSS), which incurs extra cost. Moreover, the embryo quality may deteriorate when thawed. A late severe OHSS may seriously jeopardize the woman's health, and is even life-threatening. In addition, the woman must bear the mental pressure resulted from the failure.

As in the article "What is the most relevant standard of success in assisted reproduction? The next step to improve outcomes of IVF: Consider the whole treatment" (Heijnen et al. 2004), we should define "the most informative end-point of success in IVF to be the term singleton birth rate per started IVF treatment in the *overall context of patient discomfort, complications and costs*."

In the COH cycle, the clinicians monitor the patient's individual responses based on her physiological states, such as the number and sizes of follicles, Estradiol (E_2) level, ovarian volumes, etc., and adjust dosages accordingly. On the hCG day, the clinicians want to keep the physiological states within a certain range to balance the tradeoff between the pregnancy rate and OHSS risk. However, humans tend to use myopic, experience-based policies and sometimes are influenced by the outcomes of the most recent clinical practices. As a result, the clinical decisions can therefore become subjective and inconsistent.

Our research employs a stochastic dynamic programming model to capture the stochastic and dynamic features of the COH cycle. Based on clinical guidance and records, it aims to provide a more rigorous and data-driven formalism for achieving the same goals of the clinicians.

In recent years, there have been an increasing number of operations researchers trying to apply OR methodologies to assist and improve clinical practices. For example, Alagoz et al. (2004) uses a Markov decision process (MDP) model to decide on an optimal timing of living-donor liver transplantation. They extend their work to include cadaveric livers (Alagoz et al. 2007a,b).

To the best of the authors' knowledge, this is the first research from the OR perspective on clinical decisions in IVF-ET therapy. Furthermore, this paper is unique in a sense that it incorporates dynamic clinical decisions taking into account the patients' individual stochastic responses to dosages, depicted in a correlated multi-dimensional state space. The only OR paper related to IVF-ET is by Schmittlein and Morrison (2003), where the authors use probability models to examine the two alternative payment options to infertile couples marketed by many clinics in the United States and conclude that the "money-back-guarantee program" is "too good to be true."

3 Model Formulation

We define *day 0* as the day when a patient starts exogenous gonadotropin administration in the COH cycle. Initially, clinicians make a judgment on the patient's responsiveness class, and therefore the starting gonadotropin dosage, based on her medical characteristics, such as age, BMI, number of antral follicles (follicles with 2-5 mm in diameter), ovary diameter, previous COH experiences, etc. The starting dosage is an indicator of the clinicians' initial classification of the patient's likely responsiveness/sensitivity to gonadotropin dosages.

When the diameters of the largest two follicles reach 18mm, the COH cycle ends and hCG (a gonadotropin to induce the maturity of the follicles into oocytes) is administered. The "planning horizon" (COH cycle), T, typically lasts for *6-20 days*.

In the course of the COH cycle, decisions on gonadotropin dosages are made dynamically (daily) based on each patient's physiological responses, which are dynamic and uncertain in nature. Furthermore, the growths of a patient's physiological states are correlated. The clinicians try to control these physiological states so that they fall into or are close to a target range on the hCG day, defined as a trade-off between the pregnancy rate and OHSS risk, based on clinical experiences and literature. Next, we formulate a stochastic dynamic program to model such a problem.

3.1 A Stochastic Dynamic Programming Model

We adopt the notation commonly used in the dynamic programming society as follows:

S_t = The state of the process on day t, $t = 0, ..., T$;
x_t = The decision made on day t, $t = 0, ..., T$;
W_t = Exogenous information arrives between days $t-1$ and t, $t = 1, ..., T$.

The whole process can be represented as

$$h_t = (S_0, x_0, W_1, S_1, x_1, W_2, S_2, x_2, ..., x_{t-1}, W_t, S_t).$$

3.1.1 The States

We define the states based on the patient's physiological states, resulting in a three-dimensional continuous state space.

E_t = Estradiol level (E_2, pg/ml) on day t, $E_t \in [5.0, 17000.0]$ ($\ln E_t \in [1.6, 9.7]$);
O_t = Mean diameter (mm) of the larger ovary on day t, $O_t \in [20.0, 65.0]$;
F_t = Diameter (mm) of the second largest follicle on day t, $F_t \in [3.0, 19.5]$;
S_t = (E_t, O_t, F_t), $t = 0, ..., T$.

E_2 level. Growing follicles secrete E_2. E_2 level is widely accepted as a good predictor for both the pregnancy rate and OHSS risk in the clinical literature (for example, Mathur et al. 2000; Thomas et al. 2002).

Ovarian volume. Follicles develop in the ovaries. Oyesanya et al. (1995) shows that women with moderate or severe OHSS have significantly larger mean ovarian volume on the hCG day than normal women. In practice, clinicians record ovary diameters (and use them to approximate ovary volumes). The clinicians who collaborate with us monitor the *mean ovary diameter* (average of the longest diameter and shortest diameter of the largest ovary plane) in the COH cycle and try to control it to be under 50 mm on the hCG day. To reduce the dimension of the problem, we use the mean ovary diameter of the larger ovary as a state variable. However, the model can be easily extended to include the diameters of both ovaries without affecting the structure of the problem.

The number and sizes of follicles. The number and sizes of follicles are indicators of physiological responses of the human body to gonadotropin stimulation. The retrieval of many oocytes (mature follicles) implies high OHSS risk. Al-Shawaf and Grudzinskas (2003) point out that, when more than 20 follicles are detected by ultrasound, actions should be taken to reduce OHSS risk. Another study by Asch et al. (1991) suggests that more than 30 oocytes bring 28% severe OHSS risk. On the other hand, the retrieval of too few oocytes lowers the chance of pregnancy.

We monitor the diameter of the second largest follicle because when it reaches 18 mm, the COH cycle ends. It is a continuous variable. Note that the diameter of the second largest follicle may not be exactly 18 mm on the hCG day. We also assume that the growth rate of a larger follicle is no slower than that of a smaller follicle, which is normally true except for rare cases.

It is impractical to obtain complete data on the number and sizes of all follicles in the whole COH cycle. In practice, clinicians only record the diameters of several largest follicles. In this model, we use E_t, O_t and F_t to implicitly represent the number and sizes of follicles. This makes sense as follicles secrete E_2 and they are hold in ovaries. F_t, together with E_t and O_t, somehow represents the growths of other follicles.

E_t, O_t and F_t are continuous variables. As we will discuss in Section 3.1.5, the ovary diameter and E_2 level on the hCG day are the key physiological predictors of the pregnancy rate and OHSS risk.

3.1.2 The Decision Variables

The decision variable, decision functions, and the feasible regions are defined as following.

$X^\pi(S_t)$	=	The decision function (or policy) which determines the gonadotropin dosage on day t under policy π, given current state S_t, i.e. $x_t = X^\pi(S_t)$, $t = 1, ..., T-1$;
Π	=	The set of possible policies. Each element $\pi \in \Pi$ corresponds to a different policy. $\{X^\pi(S_t)\}_{\pi\in\Pi}$ is the family of decision functions;
$\mathcal{X}_t$	=	The set of allowable decisions given the information available on day t, $t = 1, ..., T-1$.

The daily dosage decision, x_t, $t = 1, ..., T-1$, can be up to six (6) ampoules for poor responders (Hofmann et al. 1989). However, 2 and 3 ampoules are commonly adopted in clinics. (Hoomans et al. 1999, Wikland et al. (2001)). In the literature survey by Rombouts (2007), only for the poor-responsive patients does 4 ampoules initialized. Another study concerning the gonadotropin dosage for PCOS patients always suggests dosage fewer than 4 ampoules (Wely et al. 2006). In the clinic under study, the clinicians mainly use 2 and 3 ampoules and very occasionally use 4 ampoules to PCOS patients. Hence in this exploratory research, we define $\mathcal{X}_t = \{2, 3\}$, $t = 1, ..., T-1$. However, it can be easily extended to include a wider decision set.

3.1.3 The Exogenous Information Process

Patients belonging to the same responsiveness class may still respond to the gonadotropin dosage differently. The exact physiological response of a patient to a particular dosage is unknown to the clinicians when the dosage decision needs to be made. It becomes known only after the dosage is administered. The clinicians base their dosage decisions on clinical experiences and guidelines. They implicitly "optimize" the *expected* responses of patients in the responsiveness class. The stochastic responses of each class of patients can be described and governed by an underlying statistical model.

In our model, we explicitly consider each individual patient's stochastic physiological responses. We use the exogenous information process to model the stochastic part of the transition function and define

W_{t+1}	=	The exogenous information (the patient's physiological response to dosage x_t) becoming known between day t (when the dosage x_t is administered) and day $t+1$ (when the dosage decision x_{t+1} needs to be made).

3.1.4 The Transition Function

We decompose the daily change (transition) of the states into two parts: the *deterministic* part and the *stochastic* part. The deterministic part describes how the gonadotropin dosage affects the *expected* growth of follicles, ovaries, and the E_2 level. It is patient class specific. The stochastic part captures individual responses of the patients in each class. Furthermore, we can interpret the deterministic growth of the states as an direct impact of the dosage decision x_t, while the stochastic part as an exogenous random information. Hence, we find it convenient to borrow the notation of post-decision states as suggested by Powell (2007), as follows.

We model the transition function using post-decision states, and define:

$$S_t^x = S^{M,x}(S_t, x_t),$$
$$S_{t+1} = S^{M,W}(S_t^x, W_{t+1}).$$

The superscript "M" stands for "model", superscript "x" captures the impact of the decision, and "W" captures the impact of the exogenous information process.

The Deterministic Transition

We define the deterministic (post-decision) transition function using

$$S_t^x = (E_t^x, O_t^x, F_t^x), \tag{1}$$

where

$$E_t^x = E_t + \Delta_t^e(S_t, x_t),$$
$$O_t^x = O_t + \Delta_t^o(S_t, x_t),$$
$$F_t^x = F_t + \Delta_t^f(S_t, x_t).$$

and

$\Delta_t^e(S_t, x_t)$ = The deterministic (*expected*) change (growth rate) of the logarithm of E_2 level between days t and $t+1$, given state S_t and decision x_t;
$\Delta_t^o(S_t, x_t)$ = The deterministic (*expected*) change (growth rate) of the mean ovary diameter between days t and $t+1$, given state S_t and decision x_t;
$\Delta_t^f(S_t, x_t)$ = The deterministic (*expected*) change (growth rate) of the diameter of the second largest follicle between days t and $t+1$, given state S_t and decision x_t.

The Stochastic Transition

The stochastic/exogenous part of the transition function is modeled as

$$W_{t+1} = (\epsilon_{t+1}^e(S_t), \epsilon_{t+1}^o(S_t), \epsilon_{t+1}^f(S_t)), \tag{2}$$

where

$\epsilon_{t+1}^e(S_t)$ = The stochastic change of logarithm of the E_2 level between days t and $t+1$, given state S_t;
$\epsilon_{t+1}^o(S_t)$ = The stochastic change of the mean ovary diameter between days t and $t+1$, given state S_t;
$\epsilon_{t+1}^f(S_t)$ = The stochastic change of the diameter of the second largest follicle between days t and $t+1$, given state S_t.

Note that $\epsilon_{t+1}^e(S_t)$, $\epsilon_{t+1}^o(S_t)$, and $\epsilon_{t+1}^f(S_t)$ are *correlated*, which can be described using a multi-variate distribution function F. In this study, we use the trivariate truncated normal distribution, based on clinical records and expert opinions.

Combining (1) and (2), we have the complete transition function as

$$\begin{aligned} S_{t+1} &= (E_{t+1}, O_{t+1}, F_{t+1}), \\ &= S_t^x + W_{t+1}, \end{aligned} \tag{3}$$

where,

$$\begin{aligned} E_{t+1} &= E_t^x + \epsilon_{t+1}^e(S_t), \\ O_{t+1} &= O_t^x + \epsilon_{t+1}^o(S_t), \\ F_{t+1} &= F_t^x + \epsilon_{t+1}^f(S_t). \end{aligned}$$

In Chapter 9 of the book "Office-Based Infertility Practice" (Seifer and Collins 2002), it's stated that "follicular growth was noted to be *linear* during the ultrasonic examination." The *exponentially* growing pattern of E_2 level has been noticed and reported by the clinical society since decades ago (Wilson et al. 1982, Pittaway and Wentz 1983).

Consistent with the clinical literature, we observe from the clinical data that the follicles and ovaries grow *approximately linearly* while the E_2 level increases *approximately exponentially* under constant dosage. To illustrate, we plot the growths of the six largest follicles (three in the right ovary, marked with RH-1, RH-2, and RH-3, and three in the left ovary, marked with LH-1, LH-2, and LH-3), the two ovaries, and the E_2 level (in log display) of a PCOS patient under constant dosage of 2 ampoules in Figure 2.

In the COH cycle, the expected growth rates of the follicles, ovaries, and the logarithm of E_2 level are approximately linear, i.e. the growth rates do not rely on the current state. Therefore, in the current model, we approximately replace $\Delta_{t+1}^e(S_t, x_t)$, $\Delta_{t+1}^o(S_t, x_t)$, $\Delta_{t+1}^f(S_t, x_t)$ with $\Delta^e(x_t)$, $\Delta^o(x_t)$, and $\Delta^f(x_t)$, respectively. This helps us to simplify the transition function. We list the expected growth rates in the transition function in Table 1.

3.1.5 The Cost Function

In the IVF-ET therapy, the completion of the COH cycle is not the end of the story. In Figure 1, the *oocyte retrieval*, *in vitro fertilization*, and *embryo transfer* procedures all affect the pregnancy rate and OHSS risk.

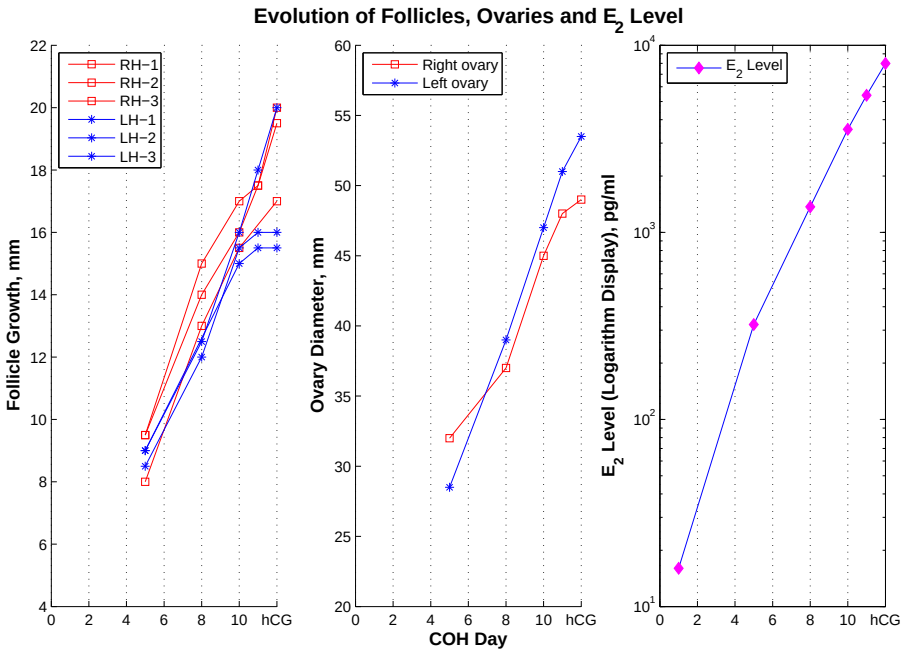


Fig. 2: A Sample Evolution of Follicles, Ovaries and E_2 Level in the COH Cycle

Table 1: Expected Growth Rates in the Transition Function

Patient class	Dosage ampoule	ln(E_2) ln(pg/ml)/day	Ovary mm/day	Follicle mm/day
Normally-R	2	0.28	1.50	1.15
	3	0.45	1.87	1.22
High-R	2	0.46	1.90	1.25
	3	0.57	2.53	1.36

Consequently, in the COH cycle, clinicians monitor and control the growths of the patient's physiological states so that they arrive at the target range of states on the hCG day. The target range is defined by incorporating the trade-off between the pregnancy rate and OHSS risk.

We define the cost function as

$$C_t(S_t, x_t) = 0, \text{ for } t = 1, ..., T-1,$$

and the cost function value $C_T(S_T)$ is evaluated on the hCG day.

In the definition of the cost function on the hCG day, $C_T(S_T)$, we do not consider F_T, because the diameter of the second largest follicle on the hCG day has little correlation with the pregnancy rate and OHSS risk. We define the target range of states in the E_T-O_T plane, as a balance of the pregnancy rate and OHSS risk, based on the clinical literature and expert opinions. We then penalize the deviations from the target range. The resulting cost function, $C_T(S_T) = f(E_T, O_T)$, is an additive piecewise linear convex function, for $S_T \in \mathcal{S}_T$, where $\mathcal{S}_T$ is the feasible region defined as $\mathcal{S}_T = \{(E_T, O_T, F_T) | E_T \in [1200.0, 17000.0], O_T \in [35.0, 65.0], F_T \in [18.0, 19.5]\}$.

The E_2 level on the hCG day is a widely accepted predictive factor for both OHSS risk and pregnancy rate. Chen et al. (2003) qualitatively describes the predictive role of E_2 level as "high E_2 levels appeared to be associated with improved treatment outcome, extremely high levels were predictive of severe ovarian hyperstimulation syndrome (OHSS)."

However, the quantitatively optimal range of E_2 level on the hCG day is still in debate in the clinical literature. Al-Shawaf and Grudzinskas (2003) believe that when the E_2 level reaches 3500 pg/ml, the gonadotropins should be withheld (a therapy called *coasting*) to reduce the OHSS risk. A more conservative strategy is to withhold gonadotropins when the E_2 level reaches 3000 pg/ml (Aboulghar and Mansour 2003). On the other hand, there are researchers arguing that coasting reduces the pregnancy rate. Peña et al. (2002) points out that relatively high E_2 level on the hCG day actually predicts high embryo transfer rate. Schmidt et al. (2004) reports that

when E2 level on the hCG day is less than 5000 pg/ml, the pregnancy increases along with the E2 level; when the E2 level is larger than 5000 pg/ml, the pregnancy rate drops. They conclude that excessive high E2 level on the hCG day is detrimental to the pregnancy rate and the "threshold estradiol level at which the pregnancy rate is reduced is 5000 pg/ml." Asch et al. (1991) points out that when the E_2 level on the hCG day is lower than 3500 pg/ml, the chance of pregnancy is significantly lower than that with the E_2 level higher than 3500 pg/ml. They also show that, if the E_2 level rises to higher than 6000 pg/ml, the severe OHSS risk increases to 38%. Clinicians cooperating with us suggest that the target (satisfactory) range of the E_2 level on the hCG day be between 3500-6000 pg/ml.

That the ovaries of patients suffering from OHSS are enlarged is commonly recognized (see Chapter 19 in "Ovulation Induction" by Tarlatzis 2002, Chapter 19 in "Manual of Ovulation Induction" by Allahbadia 2005). Oyesanya et al. (1995) shows that women with moderate or severe OHSS have significantly higher mean ovarian volume on the hCG day than normal women. The clinicians who collaborate with us try to control the mean ovary diameter on the hCG day to be under 50 mm.

Consequently, we define the target range as $E_T \in [3500.0, 6000.0]$ pg/ml and $O_T \in [45.0, 50.0]$ mm, within which we have $C_T(S_T) = 0$. The piecewise linear convex cost function on the hCG day is defined as

$$C_T(S_T) = f(E_T, O_T) = (a + b \times E_T) + (c + d \times O_T). \tag{4}$$

The parameters in (4) are listed in Tables 2 and 3, and the cost function, $C_T(S_T)$, is depicted in Figure 3. The shape of the cost function is based on the (qualitative) discussions with the clinicians. We performed an exploratory sensitivity analysis in Section 5.4.

Table 2: Parameters (in pieces) in the Cost Function: E_2 Level

E_2 (pg/ml)	1200 - 3500	3500 - 6000	6000 - 8000	8000 - 10000	10000 - 17000
a	1750.0	0.0	- 600.0	- 1400.0	- 4400.0
b	- 0.5	0.0	0.10	0.20	0.50

Table 3: Parameters (in pieces) in the Cost Function: Ovary

Ovary (mm)	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	60 - 65
c	5900.0	2700.0	0.0	-3000.0	-9600.0	-20400.0
d	-140.0	-60.0	0.0	60.0	180.0	360.0

3.1.6 The objective function

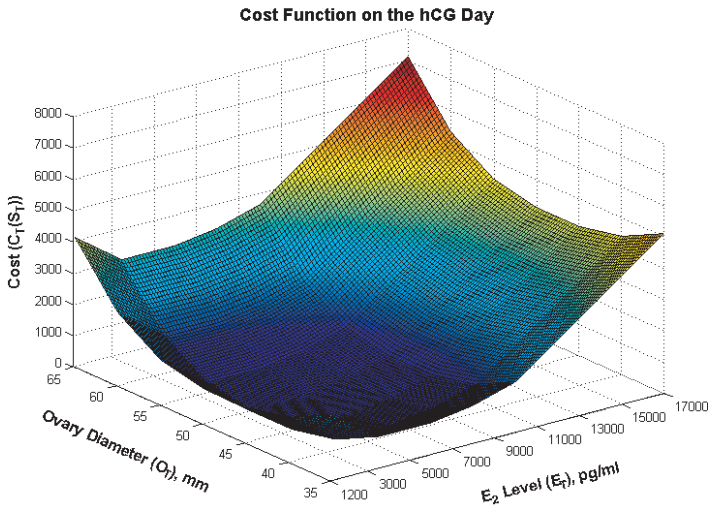
In the end, we wish to choose the best decision function (policy) to minimize the expected cost over the finite horizon (the COH cycle), as defined in (5).

$$\min_{\pi} E \sum_{t=0}^{T} C_t(S_t, X^{\pi}(S_t)). \tag{5}$$

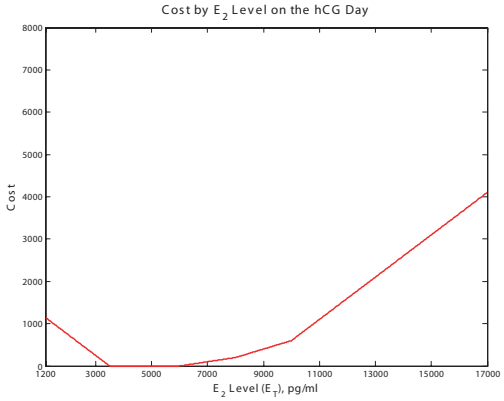
3.2 More about Classifications: Does the Distribution Count?

The patients undergoing COH are usually classified into poor, normal and high responders (Papageorgiou et al. 2002, Cai et al. 2005). As discussed in Section 2, PCOS patients are more sensitive to the gonadotropin stimulation and are therefore rarely poor responders.

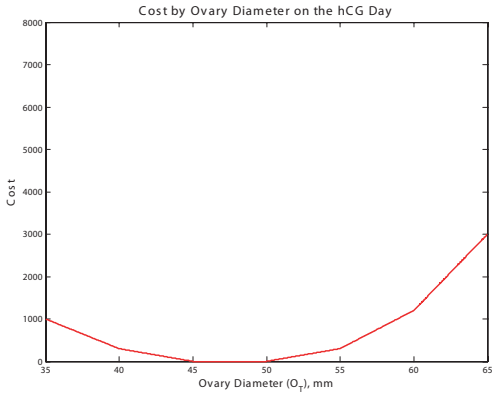
At the beginning of the COH cycle (before the observations of a patient's responsiveness), clinicians classify the patient into a likely responsiveness/sensitivity class and make decisions on the starting dosage, see Rombouts (2007). While the clinicians adjust the dosages based on observing the patient's physiological responses in the



(a) Cost Function on the hCG Day



(b) Cost by E_2 Level



(c) Cost by Ovary Diameter

Fig. 3: Cost Function on the hCG Day

COH cycle, such an initial "impression" on the patient's responsiveness will inevitably have an impact on the clinicians' dosage decisions.

When the clinicians identify a patient's responsiveness class before the COH cycle, typical *predictive* factors involve her medical characteristics such as age, BMI, previous IVF-ET experience(s), number of antral follicles (follicles with diameter 2-5 mm), and ovary diameters, the last two of which are related to the diagnosis of PCOS (Martin et al. 2006, Chang et al. 1998, Ku et al. 2006). In this study, all patients are with PCOS or PCOS potential, and all but one patient have no previous IVF-ET experiences. So *age* and *BMI* are the major medical characteristics in the initial decision of starting dosage.

In the clinic under study, a rule of thumb for patient responsiveness classification is that PCOS patients younger than 30 with BMI less than 23 or 24 are likely to be high-responsive, and therefore usually start with a dosage of 2 ampoules. On the other hand, older and heavier PCOS patients tend to be normally-responsive and usually start with a dosage of 3 ampoules. In other words, starting dosages reveal clinicians' (implicit) classification of the patient's responsiveness class.

However, an exploratory statistical analysis on the clinical records shows no significant correlations between starting dosages and the age-BMI pairs. Moreover, we observe obvious inconsistencies between the practices (decisions on starting dosage) and the claimed rule of thumb. Figure 4 illustrates the inconsistencies: quite a few young and lean patients are administered with starting dosage 3, and several old and heavy patients start with

dosage 2. Our discussions with a clinician indicate that the inconsistencies can be attributed to the subjective decisions based on their clinical experience (for example, the clinicians may tend to make more conservative decisions as a result of a recent treatment failure).

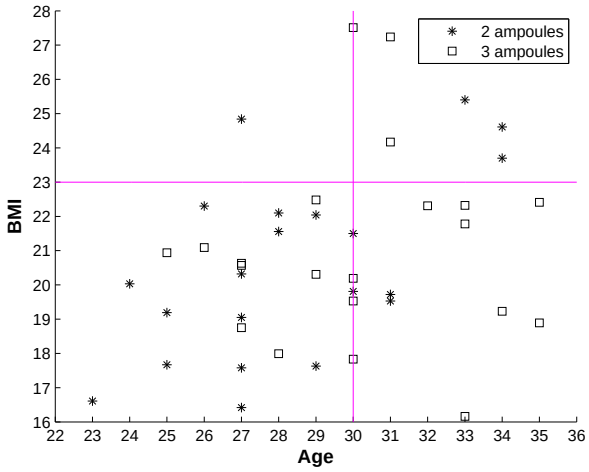


Fig. 4: Clustering Outcomes vs. Actual Classification

Each individual patient's responses to dosage are uncertain, which can be (partially) revealed by her physiological states. The key to a successful COH cycle is for the clinicians to identify the patient's likely responsiveness/sensitivity to dosage administration, and to monitor the COH cycle closely and adjust the dosages in an appropriate and timely manner. In clinical practices, an implicit assumption is that the clinicians are able to identify the patient's expected responses (responsiveness class) correctly and to make dosage decisions taking the patient's likely responsiveness into account. In the language of mathematical modeling, we assume that the underlying statistical model (distribution) governing the stochastic responses of each patient responsiveness class is known, and the clinicians make dosage decisions to optimize the expected trade-off between the pregnancy rate and OHSS risk.

Clinicians may, of course, may make an incorrect classification base on an incorrect distribution? We explore the impact of misclassification through computational experiments in Section 5.3.

4 The Solution Approach

If we discretize the state space, the problem described in Section 3 can be modeled as a finite horizon Markov decision process (MDP). This requires discretizing the set of outcomes, allowing us to compute a one-step transition matrix $\mathsf{P}(S_{t+1}|S_t, x_t)$.

The length of the treatment cycle is not deterministic, since it depends on the diameter of the second largest follicle, F_t. In addition, the cost function $C_t(S_t, x_t) = 0$ until the hCG day, at which point we incur the terminal cost $C_T(S_T)$ as defined in Section 3.1.5.

We use the (slightly modified) backward dynamic programming algorithm (Algorithm 4.1) to solve the discretized problem of (5), by solving Bellman's equations (6).

$$V_t(S_t) = \min_{x_t \in \mathcal{X}_t} \left(C_t(S_t, x_t) + \sum_{s' \in \mathcal{S}} \mathsf{P}(S_{t+1} = s'|S_t, x_t) V_{t+1}(s') \right), \quad t = 0, ..., T-1, \tag{6}$$

where $V_t(S_t)$ is the value of being in state $S_t = (E_t, O_t, F_t)$, and $\mathcal{S}$ is the discretized state space defined within $E_t \in [5.0, 17000.0]$ ($\ln E_t \in [1.6, 9.7]$), $O_t \in [20.0, 65.0]$, and $F_t \in [3.0, 19.5]$, $t = 0, ..., T$. The value of being in state S_t on the hCG day is $V_T(S_T) = C_T(S_T)$.

The transition (growth) matrix $\mathsf{P}(S_{t+1}|S_t, x_t)$ is discretized from the trivariate truncated normal distribution, given the decision x_t. Note that given S_t, the bounded state transition (growth) can eliminate infeasible states and reduce the need to search the entire state space for S_{t+1}.

Algorithm 4.1 is a (slightly modified) backward dynamic programming algorithm. For each day t from 20 to 0, we loop over all possible states S_t (E_t, O_t and F_t), and compute the value of being in this state, $V_t(S_t)$. If S_t is an end state, i.e. S_t with $F_t \geq 18mm$, we directly obtain the value of being in S_t, $V_t(S_t)$, from the cost function (4) defined in Section 3.1.5. Otherwise, we evaluate each possible decision x_t and choose the one that results in the smallest state value, $V_t(S_t)$. $V_t(S_t)$ is the smallest *expected value* in the next period, i.e., $\sum_{s' \in \mathcal{S}} \mathsf{P}(S_{t+1} = s'|S_t, x_t)V_{t+1}(s')$, where the state transition from state S_t to S_{t+1} is governed by the transition function (1), (2), and (3) defined in Section 3.1.4.

Step 0. Initialization:
  $T = 20$.
  Initialize the terminal cost $V_T(S_T) = C_T(S_T)$.
  Set $t = T - 1$.
Step 1. Do:
  If $(F_t < 18mm)$
    $V_t(S_t) = \min_{x_t} \left(\sum_{s' \in \mathcal{S}} \mathsf{P}(S_{t+1} = s'|S_t, x_t)V_{t+1}(s')\right)$, for all $S_t \in \mathcal{S}$.
  Else if $(t >= 6)$
    $V_t(S_t) = C_T(S_t)$,
    where the cost function $C_T(s)$ is defined in Section 3.1.5.
Step 2. If $t > 0$, decrement t, go to **Step 1**. Else, Stop.

Algorithm 4.1: The Backward Dynamic Programming Algorithm

The solution of (6) results in a policy/look-up table, π, under a given discretization level. Of course, the computational demands of the algorithm rise quickly when states are discretized in sufficiently fine increments. To determine the appropriate level of discretization, we compare the results using simulation.

While we use a discretized transition matrix to obtain the optimal policy under a given discretization level, in the simulation, we first simulate the state transition using the continuous transition function (trivariate truncated normal distribution), then round to the closest state, S_t, for the optimal (dosage) decision, $x_t^\pi(S_t)$, and therefore obtain the value of being in that state under policy π, $V_t^\pi(S_t)$. It is important to note that we step forward in time using the continuous state. Aggregation can produce a behavior known as *aliasing* which can destroy the Markov property (see Powell 2007, section 7.1.4), but we only use the discretized state for the purpose of evaluating the value function.

To generate random variables from the trivariate truncated normal distribution, we adopt the *normal-to-anything* Transformation (NORTA) approach (Biller and Nelson 2005). The NORTA multivariate random variable generation procedures are provided in Appendix A. Interested readers are referred to Biller and Nelson (2003, 2005); Cario and Nelson (1996, 1998) for details.

5 Computational Results

We perform three sets of experiments. First, we solve the Bellman equations (6) with different levels of discretization using the backward dynamic programming algorithm (Algorithm 4.1), and evaluate the resulting policies/look-up table with simulation. Second, we study the impact of misclassifications. Third, we perform an exploratory sensitivity analysis on the parameter settings of the cost function. Both the dynamic programming algorithm and the simulation procedure are programmed in C++. The computational experiments are performed on an Intel Core 2 Quad Q6600 CPU @ 2.40GHz and 3.25GB RAM.

5.1 Discretization

In this research, we study two classes of PCOS patients: high-responsive patients and normally-responsive patients. Each patient class has its own transition function. In the experiments, we use the same cost function, $C_T(S_T)$, for both patient classes. We experiment on seven discretization levels for each patient class. In Table 4, we list the number of (discretized) cells and the solution time required to obtain the optimal policy for each discretization level. Note that we only list the solution time of the high-responsive patient class, as the differences in solution time required for the high responsive and normally-responsive patient classes are relatively small as compared with the differences in solution time across discretization levels. We observe that the finest discretization (level 0) takes about 22 hours (78,493 seconds) to solve for the optimal policy of each patient class.

In the simulation procedure to evaluate the resulting policy, we sample 400 initial physiological states (S_0). At each initial state, we simulate 50 COH cycles, to represent the distribution of patients in the same responsiveness class with the same initial state. We use the common random stream in the simulation of all discretization levels.

We first calculate the mean and standard deviation of the cost on the hCG day of the 50 COH cycles at each initial state, and then calculate the mean and standard deviation of the cost on the hCG day of the 400 initial states, the latter of which are also listed in Table 4.

Table 4: Comparison among Discretization Levels

Discretization Level	0	1	2	3	4	5	6
Number of cells	86020	71944	57868	30636	20646	13392	6912
Solution time (seconds)	78493	54262	28998	9481	4524	1593	326
$C_T(S_T)$ (high), mean	201.2	202.5	204.8	256.1	228.6	236.8	359.8
$C_T(S_T)$ (high), stdev	175.5	173.3	174.9	169.7	174.8	179.4	240.1
$C_T(S_T)$ (normal), mean	133.9	134.1	155.9	147.6	142.3	151.1	147.3
$C_T(S_T)$ (normal), stdev	196.1	197.8	192.3	188.5	192.3	194.6	192.7

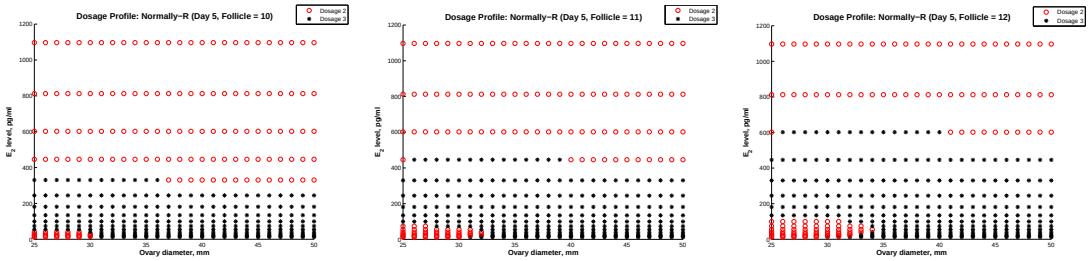
In this dynamic program, we have an ending state (S_t with $F_t \geq 18mm$) instead of an ending period. Consequently, the planning horizon is uncertain and can be as large as 20, which significantly increases the complexity of the program. After experiments on the seven discretization levels, we observe that, from level 1 to level 0, the cost function values change a little ($< 1\%$), while the solution times increase from 15 hours to 22 hours for one patient class. We decided to take level 1 as the level of discretization in the remaining experiments.

5.2 Optimal Dosage Policy

The solution to (5), i.e., the optimal dosage policy, is in the form of a lookup table where the (dosage) decision is a function of all three dimensions of the state variables. Due to the complexity of the problem structure (three-dimensional correlated states, uncertain cycle length, non-monotonic cost function on the hCG day, etc.), there are no provable monotonic control limits in the optimal dosage policy.

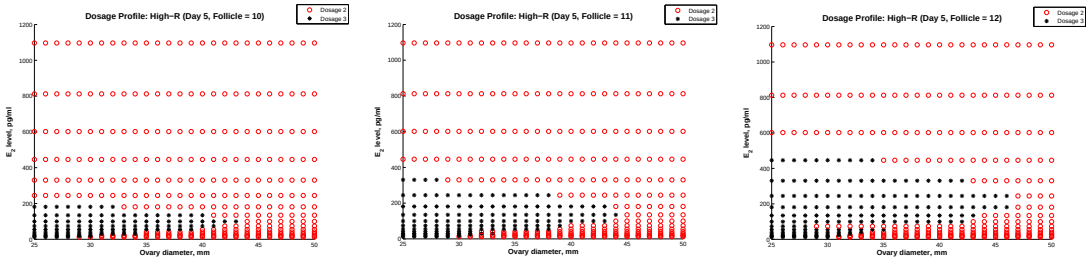
To provide an intuitive illustration of the optimal policy, we plot some sample dosage profiles of the normally and high responsive patients on different days in Figures 5 and 6. In these figures, the x-axis is ovary diameter (mm) and the y-axis is E_2 level (pg/ml). The star '*' represents dosage of 3 ampoules and the circle '$\circ$' represents dosage of 2 ampoules.

Note that in our model, we enlarged the feasible region to include *every* theoretically possible state and the resulting dosage policies (look-up tables) are complete. In the real world, the actual feasible/possible state space is much smaller. For example, it is almost impossible to have a large ovary while having a very low E_2 level, or to have a high E_2 level while having a very small ovary. Moreover, the areas around the origin (very small ovary and very low E_2) on these ovary-E_2 plots are almost impossible too, for follicle diameters $F_t = 10, 11, 12$ on day 5. As a result, the areas close to the X (ovary diameter) and Y (E_2 level) axes in Figures 5 and 6 are very unlikely in reality.



(a) Follicle Diameter = 10 mm (b) Follicle Diameter = 11 mm (c) Follicle Diameter = 12 mm

Fig. 5: Optimal Dosage Policy: Normally Responsive Patients, Day 5



(a) Follicle Diameter = 10 mm (b) Follicle Diameter = 11 mm (c) Follicle Diameter = 12 mm

Fig. 6: Optimal Dosage Policy: High Responsive Patients, Day 5

Comparing Figures 5 and 6, we observe that the areas of dosage 3 in the dosage profile for normally-responsive patients are larger than that for high responsive patients. In both patient classes, as the 2nd largest follicle becomes larger (closer to the ending state $F_t \geq 18mm$), the areas of dosage 3 become larger.

Next, we evaluate the benefits of using an optimal policy.

5.3 Impact of Patient Classification

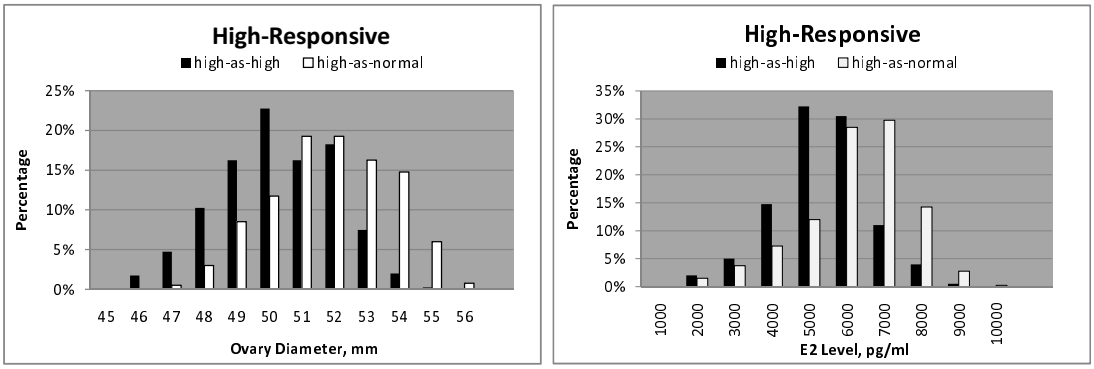
Next, we explore the answer to the question we posed in Section 3.2: what if the clinicians make a wrong classification and base their decisions on a wrong distribution?

In the following set of experiments, we explore the impact of patient misclassification. That is, we "pretend to treat" normally-responsive patients as high-responsive ones. It means that we will make dosage decisions using the policy/look-up table for high-responsive patients, while the physiological states of the patients indeed evolve under the transition functions of normally-responsive patients. Similarly, we may misclassify and "treat" high-responsive patients as normally-responsive ones.

Under discretization level 1, we sample 400 initial physiological states and 50 COH cycles at each initial state, for each patient class. We then "treat" these patients with the right and misclassified policies. Figure 7 and 9 show the distributions (histograms) of the mean ovary diameters and E_2 levels of the 400 initial physiological states on the hCG day of both patient classes. The distributions of patients on the hCG day on the E_2-ovary plane are plotted in Figures 8 and 10, respectively. In Figures 8 and 10, the inner rectangle shows the target range (E_2 level: [3500, 6000] pg/ml, ovary: [45, 50] mm), while the outer rectangle shows the ranges next to the target range in Tables 2 and 3.

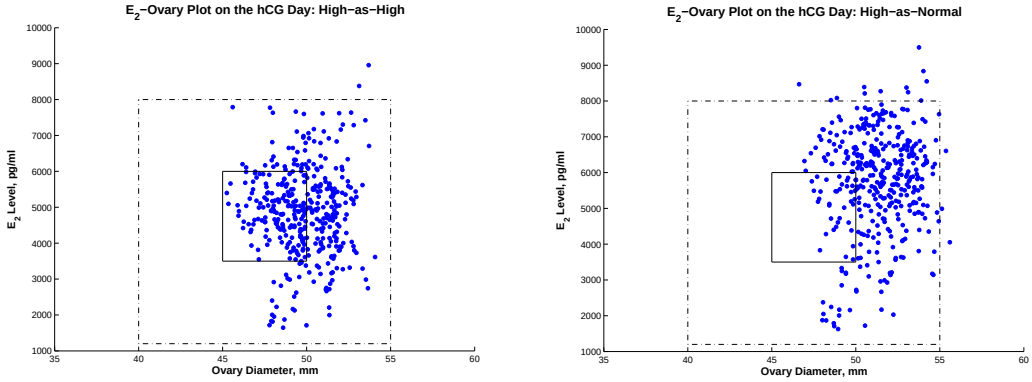
When a high-responsive patient is treated as normally-responsive, clinicians underestimate her responsiveness to gonadotropin administration and therefore tend to administer more than the necessary dosages (*over-hyperstimulation*).

The computational results show that misclassified high-responsive (high-as-normal) patients tend to have higher E_2 levels (39.8% on average) and larger ovaries (3.0% on average) over the correctly-classified (high-as-high)



(a) Ovary Diameter (b) E2 Level

Fig. 7: High-Responsive Patient Class



(a) High-as-High (b) High-as-Normal

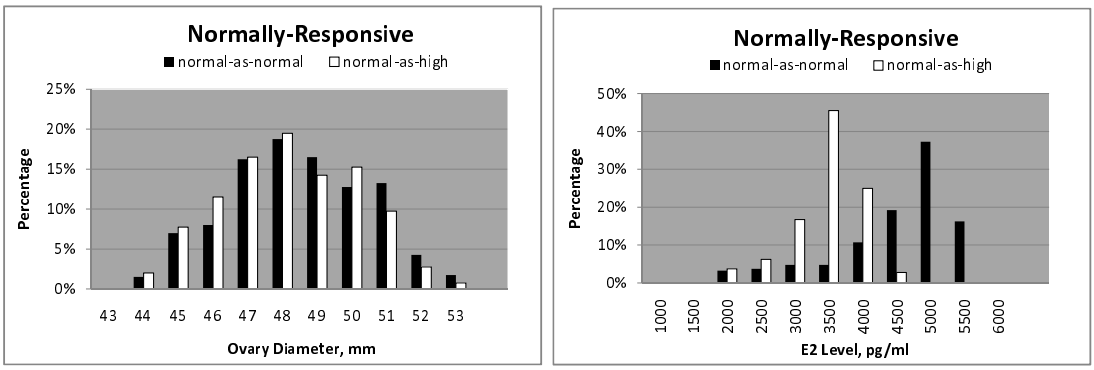
Fig. 8: E_2-Ovary Plot on the hCG Day: High Responsive

patients. Moreover, in Figure 7, we observe that the E_2 level distribution of the correctly-classified patients is more balanced around the target range ([3500, 6000] pg/ml), while that of the misclassified patients is more skewed toward the higher side. While the ovaries of the high-responsive patients tend to be large, those of the misclassified patients are more skewed toward the larger side. Compared with the correctly classified patients, 32.0% fewer of the misclassified patients have ovaries in the target range ([45, 50], mm), 28.5% fewer of the misclassified patients have E_2 level in the target range ([3500, 6000] pg/ml).

In the E_2-ovary plot on the hCG day in Figure 8, we observe that the misclassified high-responsive patients are more clustered outside the upper-right corner of the target range. 42.3% of the correctly-classified patients fall into the target range, as compared with 9.3% of the misclassified patients. Consequently, the misclassified high-responsive patients may be exposed to higher OHSS risk at the end of the COH cycle.

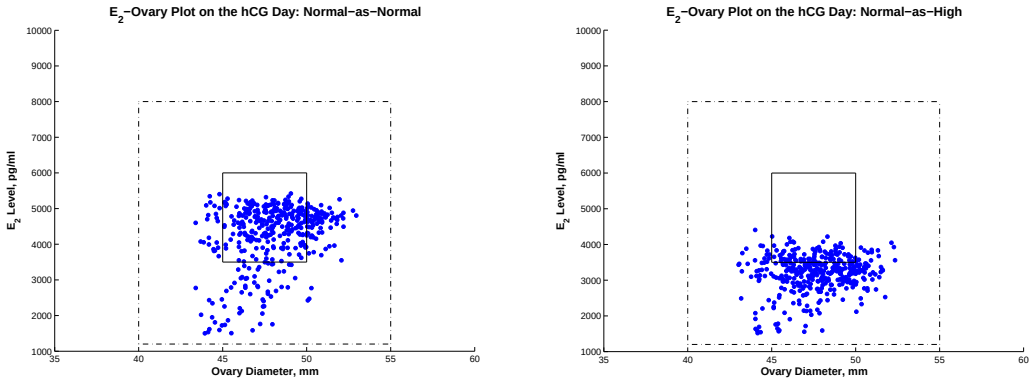
On the other hand, when a normally-responsive patient is misclassified as a high-responsive one, clinicians overestimate her responsiveness to gonadotropin administration and therefore tend to administer less than the necessary dosages (*under-hyperstimulation*).

The computational results show that the misclassified normally-responsive (normal-as-high) patients tend to have lower E_2 levels (25.1% on average), but similar (slightly smaller) ovary diameters (0.7% smaller on average), as compared with the correctly-classified (normal-as-normal) patients. In Figure 9, we observe that the distribution of E_2 levels of the misclassified patients is more skewed toward the lower side. Consequently, compared with correctly-classified patients, 6.5% fewer of the misclassified patients have ovary diameters in the target range ([45, 50] mm), and 44.5% fewer of the misclassified patients have E_2 levels in the target range ([3500, 6000] pg/ml). Moreover, 72.3% misclassified patients having E_2 levels below 3500 pg/ml. The significantly lower E_2 levels make the treatment results of the misclassified patients unsatisfactory.



(a) Ovary Diameter (b) E2 Level

Fig. 9: Normally-Responsive Patient Class



(a) Normal-as-Normal (b) Normal-as-High

Fig. 10: E_2-Ovary Plot on the hCG Day: Normally Responsive

In the E_2-ovary plot on the hCG day in Figure 10, we observe no significant difference on the ovary diameter distribution between the correctly-classified and misclassified normally responsive patients. However, the misclassified patients tend to cluster outside the lower E_2 limit of the target range. 59.5% of the correctly-classified patients fall into the target range, as compared with 21.5% of the misclassified patients. As a result, the misclassified normally-responsive patients are more likely to suffer from low pregnancy rates.

5.4 Sensitivity analysis

The shape of the cost function (4) (as depicted in Figure 3) follows clinical intuition and can be easily understood and justified qualitatively. However, the quantitative parameter settings have not been studied in the literature or analyzed in practice, as this research is the first attempt to use a stochastic dynamic programming model to formulate and evaluate the patient physiological responses to dosage and the dosage decisions in the COH cycle. In this subsection, we perform an exploratory experiment on the sensitivity analysis of the parameter settings in (4), while keeping its qualitative shape.

In this experiment, we reduce the ovary cost coefficients by 1/3 but keep the E_2 level cost coefficients in (4). Note that by doing so, we change the relative cost (penalty) of ovary and E_2 level, if the ovary or E_2 level does not fall within the target range on the hCG day. We ran the experiment (first solve for the optimal policy and then evaluate by simulation of 400 initial physiological states) under discretization level 1. In Table 5, we compare the means (with standard deviations in parentheses) of the cost, ovary diameter and E_2 level on the hCG day between the original cost function (OCF) and the modified cost function (MCF).

Table 5: Comparison between Original and Modified Cost Function

Patient class	Treatment	Cost		Ovary (mm)		E_2 level (pg/ml)	
		OCF	MCF	OCF	MCF	OCF	MCF
High-R	High	202.5 (173.3)	183.2 (164.3)	49.8 (1.8)	49.8 (1.8)	4862.9 (1207.7)	4866.1 (1197.2)
	Normal	295.8 (187.8)	261.4 (177.6)	51.3 (1.8)	51.4 (1.9)	5728.0 (1427.0)	5747.8 (1441.0)
Normally-R	Normal	134.1 (197.8)	126.4 (197.9)	48.0 (2.0)	48.0 (2.1)	4272.7 (882.3)	4280.7 (890.4)
	High	266.3 (187.6)	257.7 (185.4)	47.7 (2.0)	47.7 (2.0)	3201.5 (521.9)	3203.3 (519.3)

In Table 5, we observe significant changes on the cost on the hCG day between OCF and MCF, due to the changes of cost parameter settings. However, the changes in the cost parameter settings in (4) have a statistically negligible impact on the ovary and E_2 level on the hCG day. This result is encouraging in that it shows that the cost function is robust with respect to parameter settings while we keep the qualitative shape.

6 Conclusion

It is widely acknowledged that experience-based subjective decision making still prevails in today's clinical practices around the world. The "influence of the most recent case" decisions and the inconsistencies in clinical decisions across clinics, within the same clinic, or even by the same clinician, are not unusual. To improve patient safety and treatment effectiveness for a better health care delivery, evidence-based clinical practice and data-driven clinical decision making have been proposed (for example, Reid et al. 2005) and widely discussed recently, by both the engineering and medical societies. However, the complexity of a clinical treatment, coupled with complex human responses to the treatment, demands a good understanding of the (usually dynamic and stochastic) treatment process and the selection of appropriate modeling and algorithmic tools.

In the *controlled ovary hyperstimulation* (COH) cycle, the clinicians observe the patients' responses to the gonadotropin dosages by closely monitoring their physiological states. They then dynamically adjust dosages when necessary. Each patient's responses to dosages are uncertain when the clinicians make the dosage decisions. However, the responses of a responsiveness class of patients can be described by an underlying statistical model, based on the clinical literature, the historical clinical records, and the accumulated clinical experiences. With these statistical models in hand, we should be able to develop an optimal policy to assist the clinicians on their dosage decisions.

In this paper, we model the clinical practices in the COH treatment cycle as a stochastic dynamic program, to capture the dynamic decision process and to account for each individual patient's stochastic responses to dosage administration. We discretize the problem into a Markov decision process and solve it using a slightly modified backward dynamic programming algorithm. We then evaluate the policies using simulation, explore the impact of patient misclassification, and performed an exploratory sensitivity analysis on the cost function parameters.

The contribution of this research lies in that it is the first attempt to formulate the evolution of physiological responses under dosage administration in the COH cycle using a stochastic dynamic program. This model is used to find the optimal dosage control to achieve better and more consistent treatment outcomes. The results of this exploratory research can serve as a foundation for continuing research for a modeling/solution framework that combines statistical data analysis (data mining, learning), optimization, dynamic control, and simulation techniques to assist evidence-based, data-driven clinical decision making, which often needs to incorporate the dynamic and stochastic nature of a clinical treatment process. It provides a more rigorous clinical decision support to clinicians to hedge against possible myopic, subjective human decisions.

Acknowledgements

We are sincerely grateful to the anonymous clinicians who introduced the problem to us and guided the development of the model. We are grateful to Dr. Barry L. Nelson for introducing to us the NORTA procedure to generate the multivariate truncated normal random variable. We also thank Dr. Kaibo Wang for the discussions on the multivariate random variable generation.

References

Aboulghar, M. A. and R. T. Mansour (2003). Ovarian hyperstimulation syndrome: classifications and critical analysis of preventive measures. *Human Reproduction Update*, **9**(3), pp. 275–289.

Al-Shawaf, T. and J. G. Grudzinskas (2003). Prevention and treatment of ovarian hyperstimulation syndrome. *Best Practice & Research Clinical Obstetrics & Gynaecology*, **17**(2), pp. 249–261.

Alagoz, O., L. M. Maillart, A. J. Schaefer, and M. S. Roberts S. (2004). The optimal timing of living-donor liver transplantation. *Management Science*, **50**(10), pp. 1420–1430.

Alagoz, O., L. M. Maillart, A. J. Schaefer, and M. S. Roberts S. (2007a). Choosing among living-donor and cadaveric livers. *Management Science*, **53**(11), pp. 1702–1715.

Alagoz, O., L. M. Maillart, A. J. Schaefer, and M. S. Roberts S. (2007b). Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, **55**(1), pp. 24–36.

Allahbadia, G. N. (2005). *Manual of Ovulation Induction*. Anshan Ltd.

Asch, R. H., H.-P. Li, J. P. Balmaceda, L. N. Weckstein, and S. S. C. (1991). Severe ovarian hyperstimulation syndrome in assisted reproductive technology: definition of high risk groups. *Human Reproduction*, **6**(10), pp. 1395–1399.

Balasch, J., F. Fábregues, M. Creus, B. Puerto, J. Peñarrubia, and J. A. Vanrell (2001). Follicular development and hormone concentrations following recombinant FSH administration for anovulation associated with polycystic ovarian syndrome: prospective, randomized comparison between low-dose step-up and modified step-down regimens. *Human Reproduction*, **16**(4), pp. 652–656.

Banks, J., J. Carson, B. L. Nelson, and D. Nicol (2004). *Discrete-Event System Simulation*. Prentice Hall, fourth edition.

Barbieri, R. L. and M. D. Hornstein (1999). Assisted reproduction-In Vitro Fertilization success is improved by ovarian stimulation with exogenous gonadotropins and putuitary suppression with gonadotropin-releasing hormone analogues. *Endocrine Reviews*, **20**(3), pp. 249–252.

Biller, B. and B. L. Nelson (2003). Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation*, **13**(3), pp. 211–237.

Biller, B. and B. L. Nelson (2005). Fitting time series input processes for simulation. *Operations Research*, **53**(3), pp. 549–559.

Cai, J., H. Huang, and Y. Zhu (2005). Poor responder-high responder: The importantce of follicle stimulating hormone receptor in ovarian stimulation protocols. *Fertility and Sterility (Supplement 1)*, **84**(1), p. 44.

Cario, M. C. and B. L. Nelson (1996). Autoregressive to anything: time-series input processes for simulation. *Operations Research Letters*, **19**(2), pp. 51–58.

Cario, M. C. and B. L. Nelson (1998). Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing*, **10**(1), pp. 72–81.

Chang, M.-Y., C.-H. Chiang, T.-T. Hsieh, Y.-K. Soong, and K.-H. Hsu (1998). Use of the antral follicle count to predict the outcome of assited reproductive technologies. *Fertility and Sterility*, **69**(3), pp. 505–510.

Delvigne, A. and S. Rozenberg (2002). Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): A review. *Human Reproduction Update*, **8**(6), pp. 559–577.

Delvigne, A. and S. Rozenberg (2003). Review of clinical course and treatment of ovarian hyperstimulation syndrome(OHSS). *Human Reproduction Update*, **9**(1), pp. 77–96.

Fauser, B. C. J. M., P. Devroey, S. S. C. Yen, R. Gosden, W. F. C. Jr, D. T. Baird, and P. Bouchard (1999). Minimal ovarian stimulation for IVF: appraisal of potential benefits and drawbacks. *Human Reproduction*, **14**(11), pp. 2681–2686.

He, N. (2004). An introduction to assisted reproductive technology, part 1 (in Chinese). *China Science & Technology Education*, **6**, pp. 48–50.

Heijnen, E. M. E. W., N. S. Macklon, and B. C. J. M. Fauser (2004). What is the most relevant standard of success in assisted reproduction? The next step to improving outcomes of IVF: consider the whole treatment. *Human Reproduction*, **19**(9), pp. 1936–1938.

Hofmann, G. E., J. P. Toner, S. J. Muasher, and G. S. Jones (1989). High-dose follicle-stimulating hormone (FSH) ovarian stimulation in low-responder patients for in vitro fertilization. *Journal of in vitro fertilization and embryo transfer*, **6**(5), pp. 285–289.

Hoomans, E. H. M., A. N. Anderson, A. Loft, R. A. Leerentveld, A. A. V. Kamp, and H. Zech (1999). A prospective, randomized clinical trial comparing 150 IU recombinant follicle stimulating hormone (Puregon®) and 225 IU highly purified urinary follicle stimulation hormone (Metrodin-HP®) in a fixed-dose regimen in women undergoing ovarian stimulation. *Human Reproduction*, **14**(10), pp. 2442–2447.

Klemetti, R., T. Sevón, M. Gissler, and E. Hemminki (2005). Complications of IVF and ovulation induction. *Human Reproduction*, **20**(12), pp. 3293–3300.

Ku, S.-Y., S. D. Kim, B. C. Jee, C. S. Suh, Y. M. Choi, J. G. Kim, S. Y. Moon, and S. H. Kim (2006). Clinical efficacy of body mass index as predictor of in vitro fertilization and embryo transfer outcomes. *Journal of*

Korean Medical Science, **21**(2), pp. 300–303.

Martin, J. R., N. G. Mahutte, A. Arici, and D. Sakkas (2006). Impact of duration and dose of gonadotrophin on IVF outcomes. *Reproductive Biomedicine Online*, **13**(5), pp. 645–650.

Mathur, R. S., A. V. Akande, S. D. Keay, and J. M. Jenkins (2000). Distinction between early and late ovarian hyperstimulation syndrome. *Fertility and Sterility*, **73**(5), pp. 901–907.

Norman, R. J., D. Dewilly, R. S. Legro, and T. E. Hickey (2007). Polycystic ovary syndrome. *The Lancet*, **370**(9588), pp. 685–697.

Oyesanya, P. A., J. H. Parsons, W. P. Collins, and S. Campbell (1995). Total ovarian volume before human chorionic gonadotrophin administration for ovulation induction may predict the hyperstimulation syndrome. *Human Reproduction*, **10**(12), pp. 3211–3212.

Papageorgiou, T., J. Guibert, F. Goffinet, C. Patrat, Y. Fulla, Y. Janssens, and J.-R. Zorn (2002). Percentile curves of serum estradiol levels during controlled ovarian stimulation in 905 cycles stimulated with recombinant FSH show that high estradiol is not detrimental to IVF outcome. *Human Reproduction*, **17**(11), pp. 2846–2850.

Peña, J. E., P. L. Chang, L.-K. Chan, K. Zeitoun, M. H. Thornton, and M. V. Sauer (2002). Supraphysiological estradiol levels do not affect oocyte and embryo quality in oocyte donation cycles. *Human Reproduction*, **17**(1), pp. 83–87.

Pittaway, D. E. and A. C. Wentz (1983). Evaluation of the exponential rise of serum estradiol concentrations in human menopausal gonadotropin-induced cycles. *Fertility and Sterility*, **40**(6), pp. 763–767.

Powell, W. B. (2007). *Approximate Dynamic Programming*. John Wiley and Sons, New York.

Reid, P. P., W. D. Compton, J. H. Grossman, and G. Fanjiang (2005). *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Committee on Engineering and the Health Care System, National Academy of Engineering, Institute of Medicine. National Academies Press.

Rombouts, L. (2007). Is there a recommended maximum starting dose of FSH in IVF? *Journal of Assisted Reproduction and Genetics*, **24**(8), pp. 343–349.

Schmidt, A., S. Hahn, L. White, D. Russell, D. Kelk, and D. Smith (2004). High peak serum estradiols during IVF-ET impair pregnancy and implantation rates. *Fertility and Sterility (Supplement 2)*, **82**(2), p. 14.

Schmittlein, D. C. and D. G. Morrison (2003). A live baby or your money back: The marketing of in vitro fertilization procedures. *Management Science*, **49**(12), pp. 1617–1635.

Seifer, D. B. and R. L. Collins (2002). *Office-Based Infertility Practice*. Springer.

Tarlatzis, B. (2002). *Ovulation Induction*. ELSEVIER.

Thomas, K., T. Searle, A. Quinn, S. Wood, L. Lewis-Jones, and C. Kingsland (2002). The value of routine estradiol monitoring in assisted conception cycles. *Acta Obstetricia et Gynecologica Scandinavica*, **81**(6), pp. 551–554.

Wely, M. V., B. C. J. M. Fauser, J. S. E. Laven, M. J. Eijkemans, and F. V. D. Veen (2006). Validation of a prediction model for the follicle-stimulating hormone response dose in women with polycystic ovary syndrome. *Fertility and Sterility*, **86**(6), pp. 1710–1715.

Wikland, M., C. Bergh, K. Borg, Hillensjö., C. M. Howles, A. Knutsson, L. Nilsson, and M. Wood (2001). A prospective, randomized comparison of two starting doses of recombinant FSH in combination with cetrorelix in women undergoing ovarian stimulation for IVF/ICSI. *Human Reproduction*, **16**(8), pp. 1677–1681.

Wilson, E. A., M. J. Jawad, and T. L. Hayden (1982). Rates of exponential increase of serum estradiol concentrations in normal and human menopausal gonadotropin-induced cycles. *Fertility and Sterility*, **37**(1), pp. 46–49.

Appendix

A Procedures of Random Variable Generation from Multivariate Truncated Normal Distribution

Objective: To the generate random variable from multivariate (N) truncated normal distribution.

Precondition: We know 1) The marginal distribution of each univariate truncated normal random variable, X_i, $i = 1, ..., N$. That is, $X_i \sim$ TruncNormal $(\mu_i, \sigma_i, LB_i, UB_i)$, where μ_i and σ_i are the mean and standard deviation of the original (untruncated) normal distribution. 2) The correlation matrix Σ, with correlation between i and j denoted by $\sigma_{ij} = \text{corr}(i, j)$, $i \neq j$.

The NORTA multivariate random variable generation procedure is described in Algorithm A.1.

Step 0 Preparation (Only needs to calculate once).
  Step 0.1 For each pair of X_i & X_j, $i, j \in \{1, ..., N\}$, $i \neq j$, solve for ρ_{ij} using the procedure (Steps 1, 2, 3. Note that σ_{ij} corresponds to ρ_X.) on pages 343-344 of Banks et al. (2004).
  Step 0.2 Use ρ_{ij}, $i, j \in \{1, ..., N\}$, $i \neq j$, to construct the correlation matrix M.
  Step 0.3 Cholesky Decomposition: $M = BB^\top$.
Step 1 Generate I.I.D. RV's from standard normal distribution, Z'_i, $i = 1, ..., N$.
Step 2 Transform to tri-variate normal, $Z = BZ'$.
Step 3 For each $Z_i, i \in \{1, ..., N\}$, use the `Trunc_Normal` $(Z_i : \mu_i, \sigma_i, LB_i, UB_i)$ procedure to get X_i. Repeat Steps 1 to 3.

Algorithm A.1: NORTA Multivariate Random Variable Generation

Let $\Phi(\cdot) \equiv$ standard normal cdf, and $\Phi^{-1}(\cdot) \equiv$ standard normal inverse cdf. The `Trunc_Normal` procedure is described in Algorithm A.2.

Prerequisite These are calculated ONLY ONCE outside this procedure for each univariate truncated normal variable, X_i.
  Prereq. 1 $Z_L = (LB - \mu)/\sigma$, $Z_U = (UB - \mu)/\sigma$.
  Prereq. 2 $u_L = \Phi(Z_L)$ and $u_U = \Phi(Z_U)$.
Step 1 $u' = u_L + (u_U - u_L) \cdot u$.
Step 2 $z' = \Phi^{-1}(u')$.
Step 3 return $X = \mu + \sigma z'$.

Algorithm A.2: The Trunc_Normal $(u : \mu, \sigma, LB, UB)$ Procedure