European Journal of Operational Research xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation

Miao He [a], Lei Zhao [a,*], Warren B. Powell [b]

[a] Department of Industrial Engineering, Tsinghua University, Beijing, China
[b] Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, USA

ARTICLE INFO

Article history:
Received 8 September 2011
Accepted 29 March 2012
Available online xxxx

Keywords:
OR in health services
Approximate dynamic programming
Controlled ovarian hyperstimulation
Ovarian hyperstimulation syndrome

ABSTRACT

In the controlled ovarian hyperstimulation (COH) treatment, clinicians monitor the patients' physiological responses to gonadotropin administration to tradeoff between pregnancy probability and ovarian hyperstimulation syndrome (OHSS). We formulate the dosage control problem in the COH treatment as a stochastic dynamic program and design approximate dynamic programming (ADP) algorithms to overcome the well-known curses of dimensionality in Markov decision processes (MDP). Our numerical experiments indicate that the piecewise linear (PWL) approximation ADP algorithms can obtain policies that are very close to the one obtained by the MDP benchmark with significantly less solution time.

1. Introduction

If a couple has not conceived after trying for 1 year (or 6 months if the female is over 35), they are often considered infertile. According to the Centers for Disease Control and Prevention (CDC, http://www.cdc.gov/ART/) and the National Institute for Health and Clinical Excellence (NICE, http://www.nice.org.uk/), fertility problems affect one in six or seven couples in the United States and UK. In the United States, about 7.4% (2.1 million) of married women aging 15–44 were infertile in 2002. Infertility can be caused by women's problems, men's problems, or a mixture of both, each of which accounts for about one-third of the diagnosed infertility cases.

One necessary condition for pregnancy is that a woman must be able to release a high quality oocyte/egg. However, in a natural menstrual cycle, a woman usually produces only one oocyte/egg (in rare cases two). If the oocyte is of poor quality, the chance of pregnancy becomes very low. Even worse, some women (especially those with infertility problems) may not even be able to produce any oocytes.

Since as early as the 1950s, Robert Edwards had been working systematically to realize his vision that *in vitro fertilization* (IVF) could be useful as a treatment for infertility, which led to the world's first "test tube baby" on July 25, 1978. Since then, *assisted reproductive technology* (ART) procedures, including IVF, have been applied worldwide to help infertile couples give birth to their own children. In the United States, 148055 ART cycles were performed at 436 fertility clinics, resulting in 46326 live births and 61426 infants in 2008, as compared to 142415 ART cycles at 430 clinics, resulting in 43408 live births and 57564 infants in 2007 (Assisted Reproductive Technology Reports at the Centers for Disease Control and Prevention, http://www.cdc.gov/ART/ARTReports.htm). In Europe, 785 clinics from 29 countries reported 367066 cycles in 2004 (ESHRE, 2008). In China, about 20000 ART cycles were performed in 2004.

Among the ART approaches, IVF therapy has been the most commonly used type. 99.9% of ART procedures performed in 2008 among women using fresh nondonor eggs or embryos in the United States were IVF. In 2010, Robert Edwards was awarded the Nobel Prize in Physiology or Medicine, due to his achievement in "the development of human *in vitro fertilization* (IVF) therapy."

A typical IVF procedure (as shown in Fig. 1, He et al., 2010) begins with *controlled ovarian hyperstimulation* (COH). The COH treatment stimulates the growth of *multiple* follicles into oocytes in the woman's ovaries by administrating exogenous gonadotropin (hormone). A COH treatment cycle varies with women and may last 6–20 days (see, for example, Martin et al., 2006). At the end of the COH cycle, clinicians inject the *human chorionic gonadotropin* (hCG) to induce the final maturation of the oocytes. The oocytes are retrieved from the woman thereafter. The high quality oocytes and sperm are then placed together in a culture dish under the controlled lab environment (*in vitro fertilization*, *IVF*). If fertilization happens, the selected (high quality) embryo will be transferred into the woman's body (*embryo transfer*, *ET*). If one or more of the transferred embryos implant within the woman's uterus (positive urine test result), the cycle then may progress to clinical

* Corresponding author. Tel.: +86 10 62780815.
E-mail address: lzhao@tsinghua.edu.cn (L. Zhao).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.03.049

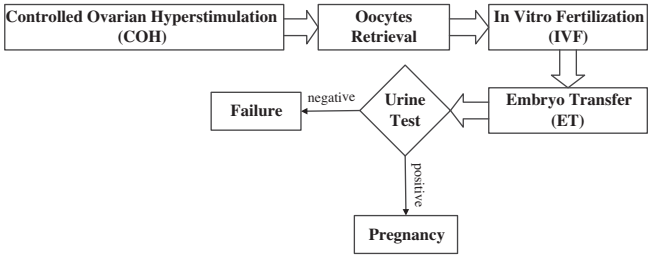


Fig. 1. A typical IVF procedure.

Table 2
The correlation matrices.

Dosage (ampoule)		State		
		$\ln(E_2)$	Ovary	Follicle
2	$\ln(E_2)$	1.00	0.56	0.58
	Ovary		1.00	0.54
	Follicle			1.00
3	$\ln(E_2)$	1.00	0.58	0.59
	Ovary		1.00	0.57
	Follicle			1.00

pregnancy. Finally, the pregnancy may lead to a live birth, the delivery of one or more live-born infants. (The birth of twins, triplets, or more is counted as one live birth.)

However, the existence of an *iatrogenic complication* named *ovarian hyperstimulation syndrome* (OHSS) casts a constant concern in the ART practices. OHSS can be mild, moderate or severe, among which severe OHSS may jeopardize the woman's life. Typical syndromes of severe OHSS include weight gain, tense ascites, hemodynamic instability (orthostatic hypotension, tachycardia), respiratory difficulty (tachypnea), progressive oliguria, and laboratory abnormalities. The severe OHSS rates reported in the medical literature are inconsistent: for example, Klemetti et al. (2005) reports that the incidence of severe OHSS varies between 0.7% and 1.7%, while Delvigne and Rozenberg (2002) report a range from 0.5% to 5%. While the pathology of OHSS has not yet been fully understood, it is widely accepted by clinicians that overdosing of exogenous gonadotropin is the main trigger of OHSS.

Moreover, one of the common causes of infertility, *polycystic ovary syndrome* (PCOS), especially exposes women undergoing the COH treatment with higher OHSS risk, because PCOS patients are *more sensitive* to gonadotropin stimulation compared with normal patients (Balasch et al., 2001; Aboulghar and Mansour, 2003; Tarlatzis, 2002). A study by Delvigne and Rozenberg (2002) 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.

With the increase of gonadotropin dosages, the probability of more follicles growing into high quality oocytes, and thus the pregnancy probability, increases. Yet the OHSS risk becomes greater with heavier dosing (over-stimulation), particularly for PCOS patients. In Heijnen et al. (2004), the authors propose "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*." Besides the overall live-birth rate, CDC's Assisted Reproductive Technology Reports present "a second measure of success based on the delivery of a live singleton. Singleton live births are a key measure of ART success because they carry a much *lower risk* than multiple-infant births for *adverse health outcomes, including prematurity, low birthweight, disability, and death*." On the other hand, the pregnancy probability becomes low with insufficient gonadotropin dosage administration (understimulation). A failed COH treatment results in a heavy burden for patients physiologically, psychologically, and economically. Therefore, in each COH treatment cycle, clinicians have to closely monitor the physiological responses of each *individual* patient, to *dynamically* decide the proper gonadotropin dosages (level of stimulation), to trade off the pregnancy probability and OHSS risk.

In He et al. (2010), we formulate the optimal dosing problem in the COH treatment cycle as a discretized Markov decision process (MDP), and solve it exactly using a slightly modified backward dynamic programming method. We then analyze the impact when clinicians misclassify a patient's physiological sensitivity to gonadotropin dosages.

However, the MDP implementation takes about 41.2 h to obtain the optimal dosing policy, as a result of the well-known *curses of dimensionality*, i.e., the explosion of the state, outcome, and decision spaces. In this paper, we design three approximate dynamic programming (ADP) algorithms to tackle the problem of dimensionality, one of which spends less than eight seconds to obtain the dosing policy with similar performance as the MDP benchmark. The successful ADP design lies in our understanding and utilization of problem structures.

Our study contributes mainly in three aspects. Firstly, we introduce the ADP modeling and algorithmic tool to assist clinical decision making (optimal dosing) in the COH treatment. Secondly, the flexibility in ADP modeling and the significant reduction of solution time offers the potential to analyze the dosage problem in more realistic and complicated settings as well as integrating successful clinical practices. Furthermore, we also experiment on some ADP algorithmic issues, which enriches the ADP literature on algorithm design and experimentation.

Next, we describe the mathematical model of the optimal dosing problem in Section 2 and briefly review the discretized MDP implementation in Section 3. In Section 4, we describe our design of ADP algorithms and key algorithmic issues. In Section 5, we report our experiment results on the ADP algorithm performances. We conclude the paper in Section 6.

2. The mathematical model

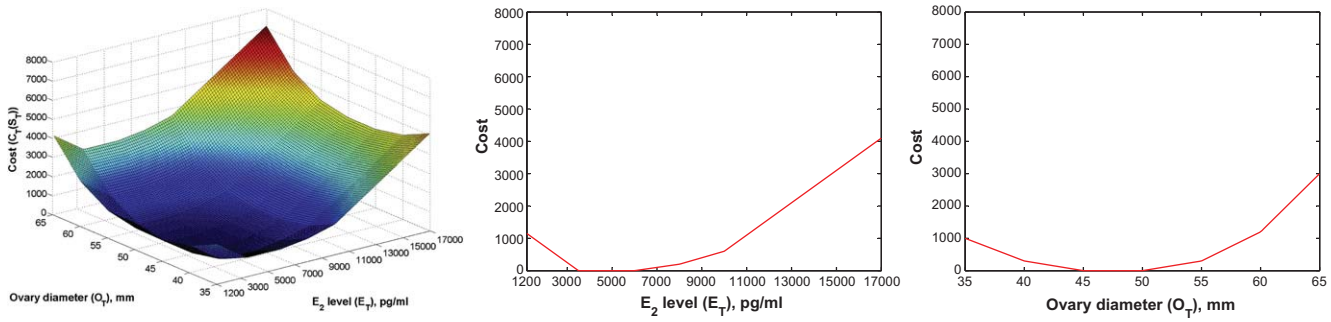
We study the *controlled ovarian hyperstimulation* (COH) treatment of high-responsive PCOS patients, who tend to be sensitive to exogenous gonadotropin administration (He et al., 2010). Each individual patient may respond differently and her physiological responses are unknown until the gonadotropin is administered. We assume the responsiveness of the PCOS patients can be described statistically, and we can obtain the probability distribution of their responsiveness based on clinical literature, statistical fitting of clinical records, and expert opinions.

In this section, we formulate a stochastic dynamic optimization model to study the dosing problem in the COH treatment cycle, where clinicians adjust the gonadotropin dosages according to

Table 1
The truncated normal distributions.

	Dosage (ampoule)	$\ln(E_2)$ (ln(pg/ml)/day)	Ovary (mm/day)	Follicle (mm/day)
μ	2	0.46	1.90	1.25
	3	0.57	2.53	1.36
σ	2	0.13	0.35	0.63
	3	0.10	0.24	0.52
LB		0.20	1.00	0.50
UB		0.60	4.00	2.00

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049



(a) Cost function on the hCG Day (b) Cost by E_2 level (c) Cost by ovary diameter

Fig. 2. Cost function on the hCG day.

the patient's stochastic physiological responses to dosage administration. The goal is to balance the tradeoff between the pregnancy probability and OHSS risk. We describe the key elements of the model as follows.

2.1. State

In the COH treatment, multiple follicles develop in the ovaries and secrete estradiol (E_2). While the *number and sizes of follicles* are indicators of human body's responses to gonadotropin administration, it is impractical to obtain complete data for all follicles. In the clinic under study, the clinicians record the diameters of a few largest follicles during the COH cycle. When the diameters of the two largest follicles reach 18 mm, clinicians induce the final maturation of the oocytes by administrating the *human chorionic gonadotropin* (hCG). This day is called the hCG day, which indicates the termination of the COH treatment cycle. *E_2 level* is a widely accepted predictor for both the pregnancy probability and OHSS risk in the clinical literature (see, for example, Aboulghar, 2003). A low E_2 level on the hCG day indicates low pregnancy probability (under-stimulation) while a high E_2 level implies high OHSS risk (over-stimulation). The sizes of ovaries reflect the number and sizes of follicles therein, and therefore indicate the patient's responses to gonadotropin stimulation. Oyesanya et al. (1995) shows that women with moderate and severe OHSS have significantly larger *ovarian volume* on the hCG day than normal women. Therefore, in this study, we use *E_2 level, mean diameter (i.e., the average of the long and short diameters) of the larger ovary, and the diameter of the second largest follicle* to represent the patient's physiological state, as follows. However, the model can be easily extended to include more dimensions in the state definition.

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:	The patient's physiological state on day t, $S_t = (E_t, O_t, F_t)$, $t = 0, \ldots, T$

At the beginning of the COH treatment cycle, the range of patients' initial physiological states is $\mathcal{S}_0 = \{S_0 | E_0 \in [5.0, 50.0], O_0 \in [20.0, 30.0], F_0 \in [3.0, 5.0]\}$. Note that in this stochastic dynamic model, we have *ending states* $\{S_T | F_T \geqslant 18.0, 6 \leqslant T \leqslant 20\}$ rather than an ending period. Therefore, the COH cycle length T is *stochastic* and ranges from 6 to 20 days.

2.2. Decision variable

In the COH treatment cycle, the dosage decision and the decision policy can be described as

$X^\pi(S_t)$:	The decision function that determines the gonadotropin dosage on day t under policy π, given state S_t, i.e., $x_t = X^\pi(S_t)$, $t = 0, \ldots, T-1$
Π:	The set of possible policies. Each element $\pi \in \Pi$ corresponds to a 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, $t = 0, \ldots, T-1$

In clinical practice, 2 or 3 ampoules are the most commonly used dosages (see, for example, Wikland et al., 2001). A study by Wely et al. (2006) suggests using dosages fewer than four ampoules for PCOS (high sensitive) patients. Therefore, we use $\mathcal{X}_t = \{2, 3\}, t = 0, \ldots, T-1$ in this exploratory study. However, the model does not put any restrictions on the feasible region of dosage decisions.

2.3. Exogenous information process

In the COH treatment cycle, the patient's physiological response (i.e., the growth of the patient's physiological state) is unknown to the clinicians when the dosage decision is made. However, based on the clinical data and clinicians' experiences accumulated in years, together with the knowledge base in the clinical literature, the stochastic physiological responses to gonadotropin administration associated with a patient sensitivity class can be described statistically.

We use a random variable W to describe the stochastic physiological response of the patient's sensitivity class. Formally, we define a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where $\mathcal{P}$ is a probability measure on the space $(\Omega, \mathcal{F})$ and the σ-algebra $\mathcal{F}$ is the set of events on the sample space Ω. We let $\omega \in \Omega$ be a sample realization from Ω. $\mathcal{F}_t \subseteq \mathcal{F}$ is defined by information available at time t. We have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, T-1$, i.e., $\mathcal{F}_t$ is a filtration. Note that the policy $\pi \in \Pi$ defined above is $\mathcal{F}_t$-measurable. Note also that the patient's stochastic responses are affected by the dosage administered. For this reason, we index the probability space by the dosage policy π and identify the probability space using $(\Omega^\pi, \mathcal{F}^\pi, \mathcal{P}^\pi)$.

Given the above definition, the exogenous information in the COH treatment cycle can be defined as

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

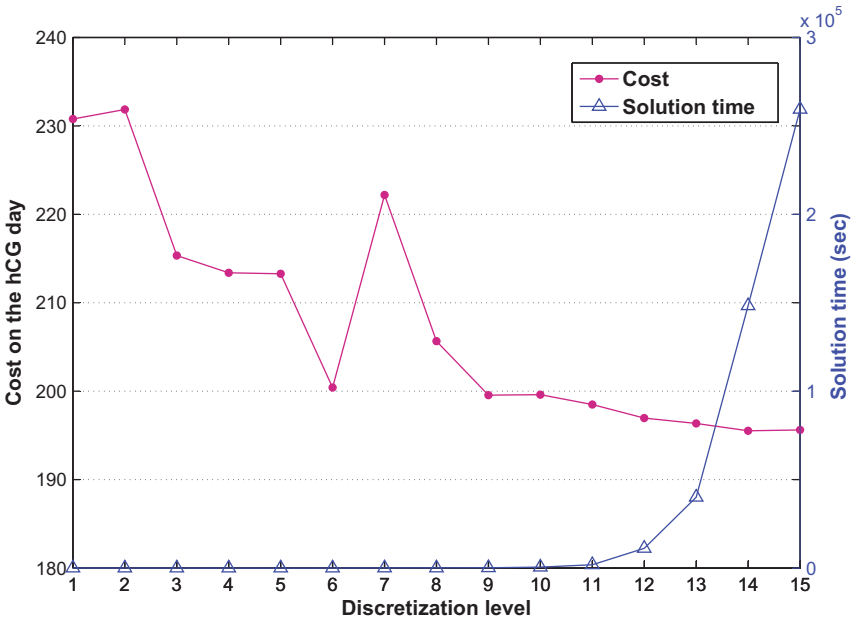


Fig. 3. MDP discretization levels: cost and solution time.

W_{t+1} : The exogenous information (the patient's stochastic 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)

$W_{t+1} = \left(W^E_{t+1}, W^O_{t+1}, W^F_{t+1}\right)$ is the information that describes the patient's physiological response given her current state S_t and dosage x_t. W_{t+1} is unknown (random) at time t, and $W_{t+1}(\omega)$ is a sample realization that we can simulate in the computer, or observe in a physical setting. Clinical literature report that, in the COH treatment cycle, the growth trajectory of the logarithm of E_2 levels, ovaries, and follicles are approximately linear (see, for example, Pittaway and Wentz, 1983; Seifer and Collins, 2002), i.e., the growth rates do not rely on the current state. Therefore, in the current model, we replace $W_{t+1}(S_t, x_t)$ with $W_{t+1}(x_t)$.

Based on the clinical records and expert opinions, we use a truncated normal distribution $\texttt{trunc\_normal}(\mu, \sigma, LB, UB)$ to describe the random growth of each state variable as in Table 1.

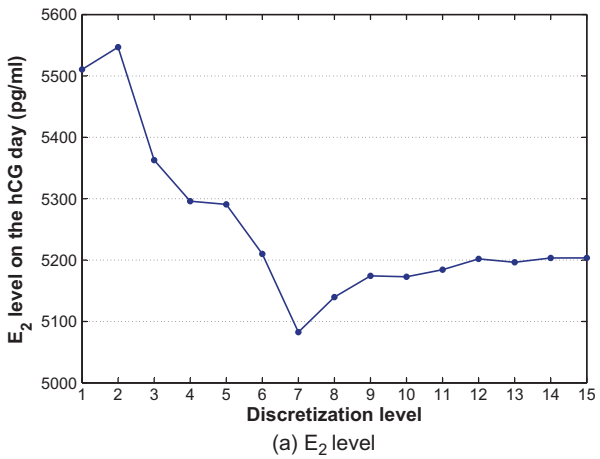
Furthermore, the follicles grow in the ovaries and growing follicles secret E_2. Therefore, the growth of the three state variables are *positively correlated* in nature. In this study, we use *trivariate truncated normal distributions*, of which the symmetric correlation matrices are defined in Table 2.

2.4. Transition function

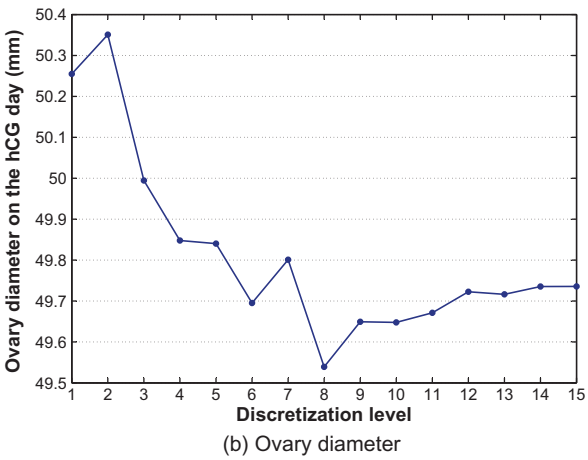
We use the transition function below to describe the growth of the patient's physiological state, which is a function of the patient's current state, the dosage decision, as well as the exogenous information that represents the patient's stochastic response to dosage administration.

$$\begin{aligned} S_{t+1} &= (E_{t+1}, O_{t+1}, F_{t+1}) = S^M(S_t, x_t, W_{t+1}) \\ &= S_t + \Delta_t(x_t) + \epsilon_{t+1}(x_t). \end{aligned} \tag{1}$$

Note that in (1) we decompose the state transition into two parts. The *deterministic* part of the transition function $\Delta_t(x_t)$ describes the *expected* growth of the patient's physiological state between days t and $t+1$, given the dosage decision x_t. The *stochastic/exogenous* part of the transition function $\epsilon_{t+1}(x_t)$ models the



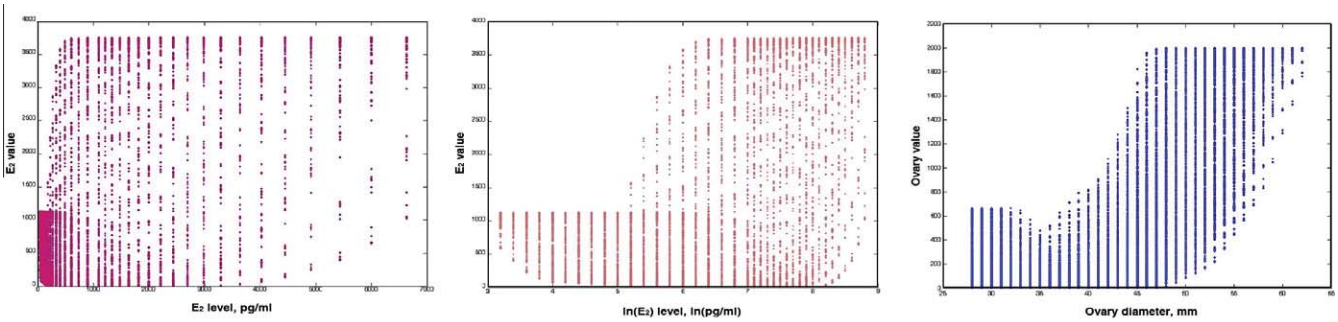
(a) E_2 level



(b) Ovary diameter

Fig. 4. MDP discretization levels: E_2 level and ovary diameter.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049



(a) $V_t^e(E_t)$ by E_2 level (b) $V_t^e(E_t)$ by $\ln(E_2)$ level (c) $V_t^o(O_t)$ by ovary diameter

Fig. 5. Value function on day 8.

stochastic growth of each individual patient's physiological state that deviates from the expected growth between days t and $t+1$, given the dosage decision x_t. The decomposition is used in the ADP algorithms around *post-decision states* (i.e., $S_t^x = S_t + \Delta_t(x_t)$) later in Section 4.1.

2.5. Cost function

As in Fig. 1, procedures after the COH cycle, such as *oocyte retrieval*, *in vitro fertilization*, and *embryo transfer*, all affect the pregnancy probability and OHSS risk. But we focus on the COH treatment cycle in this study. The outcome of a COH treatment cycle is typically measured by the E_2 level and ovary diameter on the hCG day (E_T and O_T). The target range of E_T and O_T is defined as a tradeoff between the pregnancy probability and OHSS risk. A COH treatment is considered successful if E_T and O_T are within or close to the target range. As in He et al. (2010), we define the cost function as

$$C_t(S_t, x_t) = 0, \quad \forall\, t = 0, \ldots, T-1,$$

and the cost function value $C_T(S_T)$ is evaluated on the hCG day, based on E_T and O_T. The target range is defined as $\mathcal{S}_T^{target} = \{(E_T, O_T, F_T) | E_T \in [3500.0, 6000.0], O_T \in [45.0, 50.0], F_T \geqslant 18.0\}$, within which we have $C_T(S_T) = 0$. The cost function

$$C_T(S_T) = f(E_T, O_T) = (a + b \times E_T) + (c + d \times O_T), \tag{2}$$

is a *separable additive piecewise linear (PWL) convex function* for $S_T \in \mathcal{S}_T$ (Fig. 2), where $\mathcal{S}_T$ is the feasible region defined as $\mathcal{S}_T = \{(E_T, O_T, F_T) | E_T \in [1200.0, 17000.0], O_T \in [35.0, 65.0], F_T \in [18.0, 19.5]\}$. We refer interested readers to He et al. (2010) for the parameter settings in (2) and discussions. Note that there is a cost associated with either the over- or under- stimulation case. While low pregnancy probability for the under-stimulation case is undesirable, high OHSS risk for the over-stimulation case is more prohibitive as it can become life-threatening. Therefore, we observe steeper slopes on the over-stimulation side of the cost function.

2.6. Objective function

We wish to choose the best dosage policy to minimize the expected deviation from the target state range on the hCG day. Or in other words, we wish to choose the best decision function to minimize the expected cost over the stochastic finite horizon (the COH cycle), as defined in (3).

$$\min_{\pi \in \Pi} E^{\pi} \sum_{t=0}^{T} C_t(S_t, X^{\pi}(S_t)). \tag{3}$$

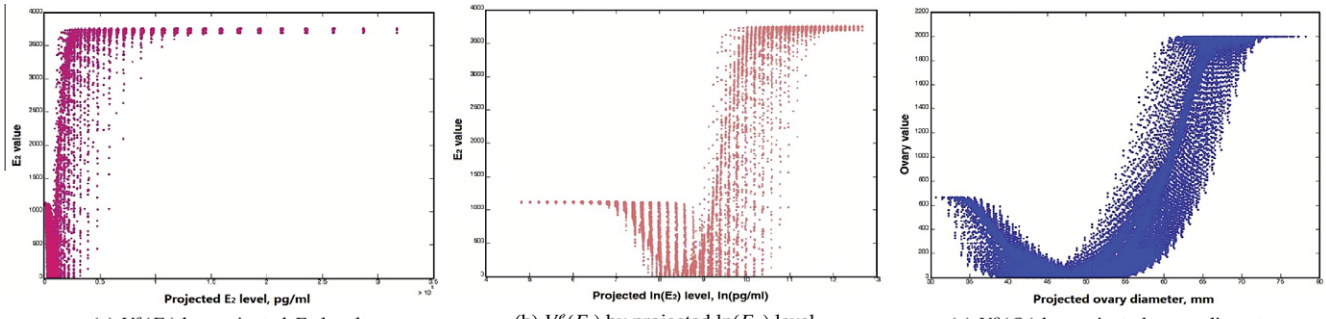
To summarize, we model the optimal dosing problem in the COH treatment cycle as a stochastic dynamic program with continuous, correlated state variables over a finite but stochastic planning horizon, the length of which is determined by the ending states ($F_T \geqslant 18$ mm). This problem is computationally intractable. In the following sections, we first briefly review the discretized MDP benchmark method (He et al., 2010), and then describe the ADP algorithms and study their performances.

3. A discretized Markov decision process (MDP) approach

If we discretize the continuous state space and transition function, the problem described in Section 2 becomes a discrete state Markov decision process (MDP) with stochastic finite planning horizon. We can solve (3) using the classical backward dynamic programming algorithm with slight modifications, by solving the Bellman equation

$$\begin{aligned} V_t(S_t) &= \min_{x_t \in \mathcal{X}_t} (c_t(S_t, x_t) + E[V_{t+1}(S_{t+1}) | S_t, x_t]), \quad t = 0, \ldots, T-1, \\ &= \min_{x_t \in \mathcal{X}_t} \left(c_t(S_t, x_t) + \sum_{s' \in \mathcal{S}_{t+1}} P(S_{t+1} = s' | S_t, x_t) V_{t+1}(s') \right), \quad t = 0, \ldots, T-1, \end{aligned} \tag{4}$$

where $V_t(S_t)$ is the value of being in state $S_t = (E_t, O_t, F_t)$ and $\mathcal{S}_{t+1}$ is the discretized state space in period $t+1$. The transition (growth) matrix $P(S_{t+1}|S_t, x_t)$ is also discretized from the transition function (1).



(a) $V_t^e(E_t)$ by projected E_2 level (b) $V_t^e(E_t)$ by projected $\ln(E_2)$ level (c) $V_t^o(O_t)$ by projected ovary diameter

Fig. 6. Value function on day 8, by projected E_t and O_t.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

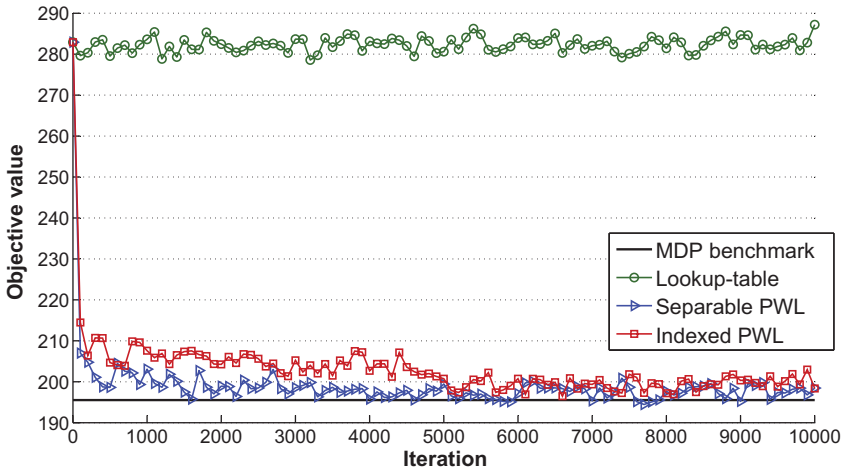


Fig. 7. Convergence plot: 10000 iterations.

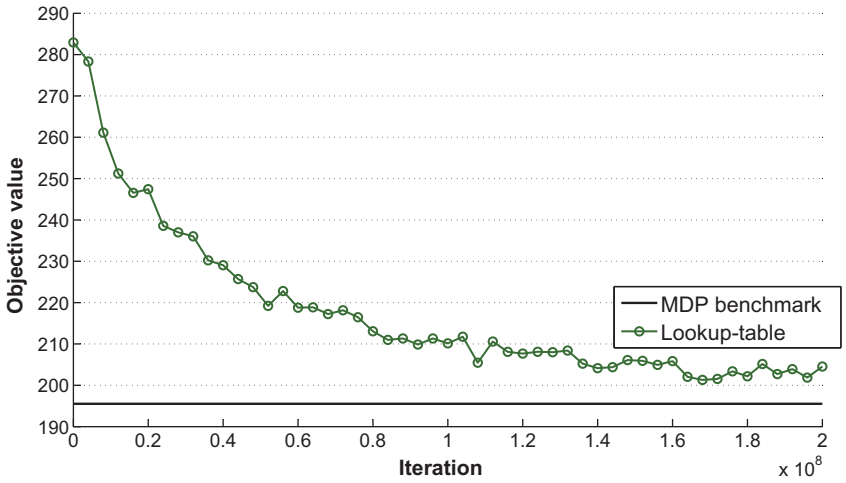


Fig. 8. Convergence plot: 200 million iterations.

Table 3
Comparison on ADP algorithms.

	Solution time	Cost on the hCG day	
		Mean	Std. dev.
MDP (level 14)	41.20 h	195.52	1.25
Lookup-table (level 14, 10000 iterations)	4.17 s	282.94	1.53
Separable PWL	7.54 s	197.40	1.28
Indexed PWL	20.10 s	200.11	1.28
Lookup-table (level 14, 200 million iterations)	12.62 h	203.77	1.30

To decide the appropriate discretization level, similar to He et al. (2010), we experiment on 15 different discretization levels, with the discretized state space ranging from about 9000 states to more than 12.5 million states. We then evaluate the resulting policies (in the form of *lookup-tables*) with simulation.

We solve (3) for an initial state range $\mathcal{S}_0$ instead of a single initial state. We randomly generate a subset of 50 initial physiological states $\mathcal{S}_{init} \subset \mathcal{S}_0$. This same set of initial states $\mathcal{S}_{init}$ is used in the simulation evaluation of the MDP policies under these 15 discretization levels for variance reduction in comparison. In each policy evaluation for each initial state, we sample 10000 paths to represent patients with the same initial state but different physiological responses. We plot the average cost on the hCG day and the solution time of each discretization level in Fig. 3, and the average E_2 levels and ovary diameter on the hCG day in Fig. 4.

As the discretization level becomes finer, the solution time grows exponentially. The average costs, E_2 levels, and ovary diameters on the hCG day appear to stabilize at discretization level 14 (in the last four discretization levels, they are within 0.74%, 0.14%, and 0.04% differences, respectively). At level 14, the state space is approximately 10 million and the solution time is approximately 41.2 h.

4. The ADP algorithms

It takes the backward dynamic programming algorithm about 41.2 h to find the optimal policy for the discretized problem with approximately 10 million states (level 14). The demanding requirements of computational resources result from the well-known curses of dimensionality in MDP and call for designing more efficient solution algorithms. Rather than solving for the value of each state exactly, approximate dynamic programming (ADP) steps forward through time via simulation and proceeds by iteratively estimating and updating the approximate value of being in a state.

We develop three versions of ADP algorithms: the lookup-table ADP, *separable* piecewise linear (PWL) function approximation and *indexed* PWL function approximation. The lookup-table representation does not rely on the problem structure, whereas the two

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

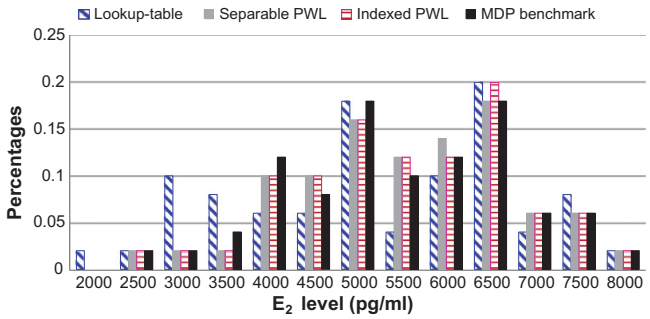


Fig. 9. Distribution of E_2 levels on the hCG day.

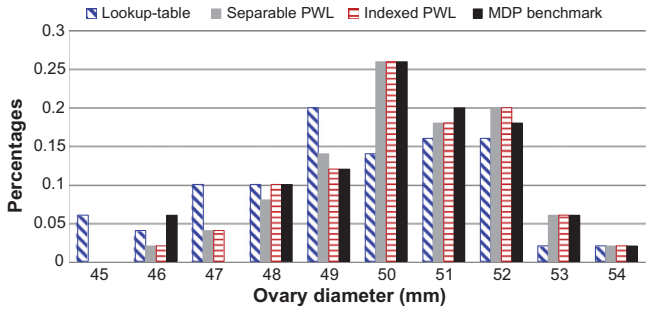


Fig. 10. Distribution of ovary diameters on the hCG day.

versions of PWL approximation methods require a certain level of understanding on the problem characteristics.

4.1. Key algorithmic issues in ADP

In this subsection, we discuss the key algorithmic issues in the design of ADP algorithms to balance the computational effort and solution accuracy and to effectively solve the stochastic dynamic problem (3) as described in Section 2.

4.1.1. Post-decision state

Solving the Bellman equation (4) involves computing the expectation within the min operator, which suffers from the curse of dimensionality in the outcome space. In approximate dynamic programming, we introduce the *post-decision state*, defined as the state of the system after we have made a decision but before any new information has arrived. In this problem, the post-decision state can be expressed as $S_t^x = S^{M,x}(S_t, x_t) = S_t + \Delta_t(x_t)$. Note that S_t is the state immediately before we make a decision, sometimes denoted as the *pre-decision state*. That is, before the clinicians make their dosage decision, they can observe the patient's physiological state, S_t. After the clinicians have made the dosage decision, they only know the *expected* growth of the patient's physiological state, but do not know her exact response. When the patient's response is observed, the next pre-decision state becomes $S_{t+1} = S_t^x + \epsilon_{t+1}(x_t) = S_t + \Delta_t(x_t) + \epsilon_{t+1}(x_t)$. There is a vast literature in reinforcement learning that uses Q-learning, which involves estimating the value of a state-action pair. Technically, a state-action pair is a form of post-decision state (Powell, 2007, p. 144), but our formulation of a post-decision state is more compact. When making decisions, we use

$$\hat{v}_t^n(S_t^n) = \min_{x_t \in \mathcal{X}_t}\left(c_t(S_t^n, x_t) + \overline{V}_t^{n-1}(S_t^{x,n})\right), \quad t = 0, \ldots, T-1, \tag{5}$$

where we replace the expectation of the exact value function around the pre-decision state $E[V_{t+1}(S_{t+1})]$ in (4) with the approximate value function around the post-decision state $\overline{V}_t^{n-1}(S_t^{x,n})$. The superscript n is the iteration counter and $\hat{v}_t^n$ is the new observed value for the visited state S_t^n.

Note that $\hat{v}_t^n(S_t^n)$ is a sample of the value of being in state S_t^n, and it is also a sample of the value that put us in state $S_{t-1}^{x,n}$, since the transition from $S_{t-1}^{x,n}$ to S_t^n requires only the realization of random exogenous information. Therefore, we can update the estimate of state value $\overline{V}_{t-1}^n(S_{t-1}^{x,n})$ with the new observations $\hat{v}_t^n(S_t^n)$, as

$$\overline{V}_{t-1}^n(S_{t-1}^{x,n}) = (1-\alpha_{n-1})\overline{V}_{t-1}^{n-1}(S_{t-1}^{x,n}) + \alpha_{n-1}\hat{v}_t^n(S_t^n), \tag{6}$$

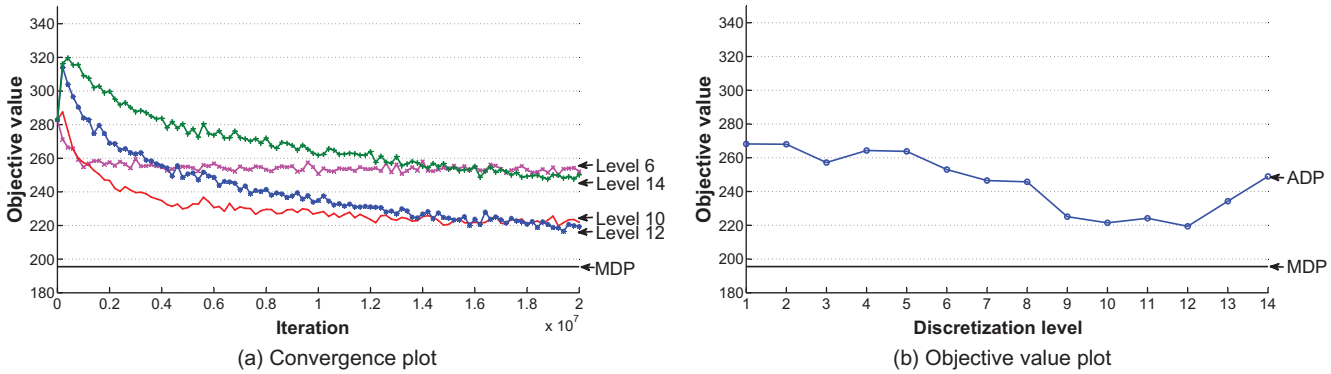
where α_{n-1} is a smoothing factor (*stepsize*) governed by a particular stepsize rule.

4.1.2. Stepsize rules

The choice of stepsize rule has a significant impact on the convergence behavior of an ADP algorithm and the quality of the solution. *Deterministic* stepsize rules do not change with the observed data in the process of approximating state values, while *stochastic* stepsize rules adapt to the actual trajectory of the algorithm. George and Powell (2006) propose an optimal stochastic stepsize rule for nonstationary data, i.e., the *Bias-adjusted Kalman Filter* (BAKF) stepsize, which is given by

$$\alpha_{n-1} = 1 - \frac{(\bar{\sigma}^2)^n}{(1+\bar{\lambda}^{n-1})(\bar{\sigma}^2)^n + (\bar{\beta}^n)^2}, \tag{7}$$

where $\bar{\lambda}^n$ is computed recursively using



(a) Convergence plot (b) Objective value plot

Fig. 11. Lookup-table ADP: comparison on discretization levels.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

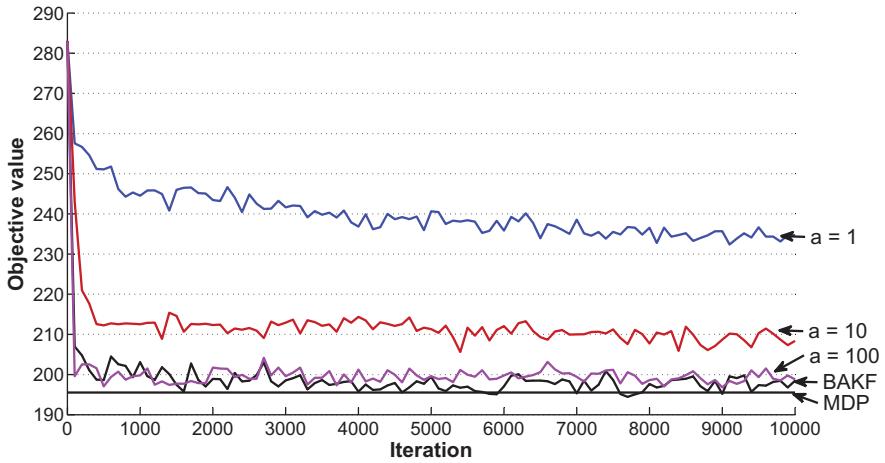


Fig. 12. Stepsize rules: BAKF vs. harmonic.

Table 4
Distribution of E_2 levels and ovary diameters on the hCG day.

	E_2 levels			Ovary diameters		
	Below (%)	In target range (%)	Above (%)	Below	In target range (%)	Above (%)
MDP (level 14)	4	64	32	–	54	46
Lookup-table (level 14, 10000 iterations)	14	52	34	–	64	36
Separable PWL	4	64	32	–	54	46
Indexed PWL	4	62	34	–	54	46

$$\bar{\lambda}^n = \begin{cases} (\alpha_{n-1})^2, & n = 1, \\ (1-\alpha_{n-1})^2 \bar{\lambda}^{n-1} + (\alpha_{n-1})^2, & n > 1. \end{cases} \tag{8}$$

$(\bar{\sigma}^2)^n$ denotes the estimate of the variance of the value function $\overline{V}_t^n(S_t^n)$ and $\bar{\beta}^n$ denotes the estimate of bias due to smoothing a non-stationary data series. The BAKF stepsize rule balances the estimate of the noise $(\bar{\sigma}^2)^n$ as well as the estimate of the bias $\bar{\beta}^n$ due to the transient nature of the data in the ADP solution process. Interested readers may refer to Section 6.5.3 in Powell (2007) and George and Powell (2006) for more details.

To compare, we also experiment with a deterministic stepsize rule, i.e., the harmonic stepsize rule, as in (9),

$$\alpha_{n-1} = \frac{a}{a+n-1}, \tag{9}$$

where a is a constant.

4.1.3. An initial state range

At the beginning of the COH treatment cycle, the initial physiological state varies with patient. We therefore have an initial range of states rather than a single initial state. In each iteration of the ADP algorithm, we randomly generate an initial state ($S_0 \in \mathcal{S}_0$) that represents a particular patient from the same responsiveness class, and start from that state to simulate the COH cycle and update the estimated state values.

4.2. Lookup-table ADP

Recall that the problem has continuous state and outcome spaces. The lookup-table ADP requires discretization of the state space as in MDP. However, instead of solving (4) exactly around pre-decision states as in Section 3, we approximate the value of being in a *post-decision* state S_t^x, $\overline{V}_t^n(S_t^x)$, via simulation and iterative update. Algorithm 4.1 describes the lookup-table ADP algorithm around post-decision states.

Algorithm 4.1. The double-pass lookup-table ADP algorithm

Step 1. Initialize $\overline{V}_t^0(S_t^x), \forall\, S_t^x, \forall\, t$. Set $n = 1$.
Step 2. Iterative updates.
  Step 2.1. Randomly generate an initial state $S_0^n \in \mathcal{S}_0$.
  Step 2.2. Generate a sample path ω^n.
Step 3. The *forward* pass. Set $t = 0$.
  Step 3.1. For S_t^n, solve for $x_t^* = \arg\min_{x_t \in \mathcal{X}_t} \overline{V}_t^{n-1}(S_t^{x,n})$.
  With probability $1-\rho$, set $x_t^n = x_t^*$ (exploitation).
  With probability ρ, choose a decision x_t^n from $\mathcal{X}_t \setminus \{x_t^*\}$ (exploration).
  Step 3.2. Observe the state transition and record the visited post-decision states $S_t^{x,n}$.

$$S_t^{x,n} = S^{M,x}(S_t^n, x_t^n) = S_t^n + \Delta(x_t^n),$$
$$S_{t+1}^n = S^M(S_t^n, x_t^n, W_{t+1}(\omega^n)) = S_t^n + \Delta(x_t^n) + \epsilon(\omega_{t+1}^n),$$

  where $\Delta(x_t^n)$ is the expected growth, and $\epsilon(\omega_{t+1}^n)$ is the random deviation from $\Delta(x_t^n)$.
  Step 3.3. Ending state check.
  If $F_{t+1} \geqslant 18mm$,
  mark day $t+1$ as the hCG day, i.e., $T = t+1$,
  compute $C_T(S_T^n)$, and go to **Step 4**;
  else
  increment t and go to **Step 3.1**.
Step 4. The *backward* pass. For $t = T-1, \ldots, 0$, do
  For the visited post-decision state, $S_t^{x,n}$, update its value function approximation.

$$\overline{V}_t^n(S_t^{x,n}) = (1-\alpha_{n-1})\overline{V}_t^{n-1}(S_t^{x,n}) + \alpha_{n-1} C_T(S_T^n).$$

Step 5. Let $n = n+1$. If $n \leqslant N$, go to **step 2**.
Step 6. Return the value function (lookup-table) $(\overline{V}_t^N)_{t=0}^T$.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

We use the double-pass algorithm (Chapter 8 of Powell, 2007), which corresponds to TD (1) in *temporal difference learning* (Sutton and Barto, 1998). The double-pass ADP algorithm uses the accumulated path costs $\sum_{i=t}^{T} C_i(S_i^n)$ to update the value of the visited post-decision state on period $t-1$. In our case, we use the cost on the hCG day, $C_T(S_T)$, to update the estimated value of the visited post-decision state, since zero costs are incurred before the hGG day. In the forward pass (Step 3), we simulate the COH treatment process until the hCG day (i.e., when $F_T \geqslant 18mm$) and evaluate the treatment cost $C_T(S_T)$ using (2). In the backward pass (Step 4), we update the value function approximation of the (post-decision) states visited in the forward pass with $C_T(S_T)$,

$$\overline{V}_t^n(S_t^{x,n}) = (1-\alpha_{n-1})\overline{V}_t^{n-1}(S_t^{x,n}) + \alpha_{n-1} C_T(S_T^n). \tag{10}$$

Note that we have $C_t(S_t) = 0$, $t = 0,\ldots,T-1$, in this problem.

A fundamental challenge with approximate dynamic programming is the "exploration vs. exploitation" problem. Do we make a decision to *explore* the value of being in a state? Or do we *exploit* our current estimates of downstream state values to make the best possible decision based on our current belief? It is quite easy to be stuck in a local solution simply because of poor value estimates of being in some states in pure exploitation, while convergence can be extremely slow in pure exploration. In this paper, we employ a simple *mixed exploration and exploitation strategy* (Chapter 10 of Powell, 2007) with a fixed per period exploration rate ρ. As in Step 3.1, with probability ρ, we do not use the "optimal" dosage decision based on the current value function approximation $\overline{V}_t^{n-1}$; instead, we choose an alternative dosage decision at random simply to learn its value. A too large value of ρ will slow down the convergence, but a too small value of ρ may deteriorate the effect of exploration. In approximate dynamic programming, there are still many unanswered questions on understanding how well we know the value of being in a state and how to make a decision to better help us learn this value.

4.3. Separable piecewise linear value function approximation

While the lookup-table ADP has the advantage of being independent of problem structures, it struggles with the large state space. Moreover, the algorithm efficiency is relatively low, because the algorithm updates only the visited states. With a proper design of the *value function* approximation based on the structural properties of state variables, we can improve the algorithm performance with more efficient value updates.

4.3.1. Value function approximation

Note that we evaluate the effect of dosage policy in the COH treatment cycle by the E_2 level (E_T) and ovary diameter (O_T) on the hCG day, when the diameter of the second largest follicle reaches the diameter of 18 mm ($F_T \geqslant 18$). In this study, the cost function on the hCG day $C_T(S_T)$ is separable and additive in E_T and O_T, as in (2).

In MDP, the state value $V_t(S_t)$, $t = T-1,\ldots,0$, can be calculated separately with regard to the E_2 level $(V_t^e(E_t))$ and ovary diameter $(V_t^o(O_t))$, in a backward manner. That is, $V_t(S_t) = V_t^e(E_t) + V_t^o(O_t)$. The impact of another state, F_t, is implicitly reflected in the length of the treatment cycle and accounted for in the expectation calculation at each state, as in (4).

However, in ADP, simply approximating $V_t(S_t)$ by E_2 level and ovary diameter, i.e., $\overline{V}_t(S_t) = \overline{V}_t^e(E_t) + \overline{V}_t^o(O_t)$, results in a poor value approximation, because we ignore the impact of the follicle diameter (F_t). More specifically, the same E_2 level may have different values, depending on the sizes of the follicle diameter, which indicates the closeness to the hCG day. This is illustrated in Fig. 5, where we plot the state values on day 8 with regard to E_t and O_t, based on the MDP data.

To capture the impact of F_t on $\overline{V}_t^e(E_t)$ and $\overline{V}_t^o(O_t)$, we first use F_t to estimate how many days remaining until the hCG day and then estimate (project) the E_2 level and ovary diameter on the hCG day, based on E_t, O_t and their growth rate estimates. We define the *projected E_2 level* and *ovary diameter* on the hCG day, based on $S_t = (E_t, O_t, F_t)$, as

$$P_t^e(S_t) = exp\left\{E_t + \overline{G}_t^e \times \frac{(\overline{F}_T - F_t)}{\overline{G}_t^f}\right\},$$

and

$$P_t^o(S_t) = O_t + \overline{G}_t^o \times \frac{(\overline{F}_T - F_t)}{\overline{G}_t^f},$$

where

$\overline{G}_t^e$:	an estimate of the growth rate of the $\ln(E_2)$ level on day t
$\overline{G}_t^o$:	an estimate of the growth rate of the ovary diameter on day t
$\overline{G}_t^f$:	an estimate of the growth rate of the follicle diameter on day t
$\overline{F}_T$:	the average follicle diameter on the hCG day

In Fig. 6, we plot the state values on day 8 with regard to the projected E_2 levels and ovary diameters on the hCG day, where we observe cleaner shapes as compared to Fig. 5. We will address the impact of the correlation between the (projected) E_2 levels and ovary diameters using an indexed piecewise linear function approximation in Section 4.4.

For notation simplicity, we drop the S_t and denote $P_t^e(S_t)$ and $P_t^o(S_t)$ as P_t^e and P_t^o in the rest of the paper. On day t, we define the *value function* with regard to P_t^e and P_t^o as

$$\overline{V}_t(S_t) = \overline{V}_t^e(P_t^e) + \overline{V}_t^o(P_t^o), \tag{11}$$

where

$\overline{V}_t(S_t)$:	the approximate value of state S_t on day t
$\overline{V}_t^e(P_t^e)$:	the approximate value with regard to P_t^e (projected E_2 level on the hCG day)
$\overline{V}_t^o(P_t^o)$:	the approximate value with regard to P_t^o (projected ovary diameter on the hCG day)

The value functions (as in Fig. 6) are convex in the middle and concave in two tails. This shape can be intuitively understood: too high value of the projected E_2 level or ovary diameter indicates possible over-stimulation and too low value indicates possible under-stimulation. The adverse effect increases when further deviating from the "right" physiological state range. The two tails account for the boundary effect of the state range.

The shapes of the value functions around post-decision states are similar to those around pre-decision states. We approximate each component in the post-decision state value function $(\overline{V}_t^e(P_t^e)$ and $\overline{V}_t^o(P_t^o))$ using a *piecewise linear function*. Note that we solely use value functions around post-decision states in this study, so we drop the superscript "x" in the following text for notational simplicity. With the convexity/concavity properties, we are able to update the value function using the SPAR algorithm (Section 11.3 of Powell, 2007).

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

4.3.2. Separable piecewise linear function approximation

Algorithm 4.2 summarizes the separable piecewise linear (PWL) value function approximation ADP algorithm. We define

$\hat{\psi}^{e,n}_{t,j}(P^e_t)$:	the *observed* slope of segment j in $\overline{V}^{e,n}_t(P^e_t)$ in iteration n
$\hat{\psi}^{o,n}_{t,j}(P^o_t)$:	the *observed* slope of segment j in $\overline{V}^{o,n}_t(P^o_t)$ in iteration n
$\bar{\psi}^{e,n}_{t,j}(P^e_t)$:	the *smoothed* slope of segment j in $\overline{V}^{e,n}_t(P^e_t)$ after n iterations
$\bar{\psi}^{o,n}_{t,j}(P^o_t)$:	the *smoothed* slope of segment j in $\overline{V}^{o,n}_t(P^o_t)$ after n iterations
N:	the maximum number of iterations before terminating the algorithm

Note that a segment (piece) of the piecewise linear function is a discretization interval here. That is, we use the same discretization level in the PWL value function approximation algorithms as in the lookup-table ADP (no aggregation is needed). The algorithmic efficiency gains mainly from the efficient update of the PWL value functions due to their convexity/concavity. We set the value of the smallest point of the feasible state range to be a fixed value. Given the set of slopes for all the segments, we can derive the value function $\overline{V}$. Similar to the lookup-table algorithm, we use the double-pass algorithm in the value function update. In the forward pass, we obtain the post-decision state value from a separable piecewise linear function instead of looking up in a table as in Algorithm 4.1. In the backward pass, we perturb the visited post-decision state and obtain the numerical derivative of state values in order to update the slope of the visited segment in the value function. We perturb the E_2 level by δ^e to obtain a perturbed E_2 level, $\widetilde{P}^e_t$. Denote the value for the perturbed E_2 level as $\tilde{v}^{e,n}_t\left(\widetilde{P}^e_t\right)$. Similarly, we obtain $\delta^o, \widetilde{P}^o_t$, and $\tilde{v}^{e,n}_t\left(\widetilde{P}^e_t\right)$ for the ovary diameter. The new observed slopes for the E_2 value function and ovary value function become

$$\hat{\psi}^{e,n}_{t,j}(P^e_t) = \begin{cases} \left[\hat{v}^{e,n}_t(P^e_t) - \tilde{v}^{e,n}_t\left(\widetilde{P}^e_t\right)\right]\Big/\delta^e, & \text{if we visit segment } j, \\ \bar{\psi}^{e,n-1}_{t,j}(P^e_t), & \text{o.w.} \end{cases} \tag{12}$$

and

$$\hat{\psi}^{o,n}_{t,j}(P^o_t) = \begin{cases} \left[\hat{v}^{o,n}_t(P^o_t) - \tilde{v}^{o,n}_t\left(\widetilde{P}^o_t\right)\right]\Big/\delta^o, & \text{if we visit segment } j, \\ \bar{\psi}^{o,n-1}_{t,j}(P^o_t), & \text{o.w.} \end{cases} \tag{13}$$

With the new observed slopes, we smooth the estimated slopes with

$$\bar{\psi}^{e,n}_{t,j}(P^e_t) = (1-\alpha_{n-1})\bar{\psi}^{e,n-1}_{t,j}(P^e_t) + \alpha_{n-1}\hat{\psi}^{e,n}_{t,j}(P^e_t), \tag{14}$$

and

$$\bar{\psi}^{o,n}_{t,j}(P^o_t) = (1-\alpha_{n-1})\bar{\psi}^{o,n-1}_{t,j}(P^o_t) + \alpha_{n-1}\hat{\psi}^{o,n}_{t,j}(P^o_t). \tag{15}$$

If any violations to concavity or convexity occur, we follow the SPAR (the Separable, Projective Approximation Routine) algorithm to maintain those properties (Step 4.5).

Assume that our current value function approximation $\overline{V}^{n-1}$ is in the form of a set of slopes, $\left\{\bar{\psi}^{n-1}_j\right\}$ (Note that for simplicity, we drop the superscript e or o and the subscript t here). Assume we have visited the segment j^n and observed an estimate of the slope $\hat{\psi}^n$, and we use the SPAR algorithm to maintain the concavity of $\overline{V}^n$. Let $\bar{y}^n$ temporarily store the updated set of slopes, that is,

$$\bar{y}^n_j = \begin{cases} (1-\alpha_{n-1})\bar{\psi}^{n-1}_j + \alpha_{n-1}\hat{\psi}^n, & j = j^n, \\ \bar{\psi}^{n-1}_j, & \text{otherwise.} \end{cases} \tag{16}$$

If $\bar{y}_j \geqslant \bar{y}_{j+1}$, then the function $\overline{V}^n$ maintains concavity. Otherwise, then either $\bar{y}_{j^n} < \bar{y}_{j^n+1}$ or $\bar{y}_{j^n-1} < \bar{y}_{j^n}$. We fix this violation by solving the following projection problem.

$$\min_v \|v - \bar{y}^n\|^2, \tag{17}$$

$$\text{s.t.} \quad v_j - v_{j+1} \geqslant 0. \tag{18}$$

The projection problem can be solved easily. For example, if we observe a violation to the left ($\bar{y}_{j^n-1} < \bar{y}_{j^n}$) after the update, we can average the updated segment (j^n) with all the segments to the left that create the violation.

Algorithm 4.2. The double-pass separable PWL ADP algorithm

Step 1. Initialize.
  Step 1.1. Set initial estimates of the slopes $\bar{\psi}^{e,0}_{t,j}(P^e_t)$ and $\bar{\psi}^{o,0}_{t,j}(P^o_t)$ to zero for each segment j of the post-decision state value functions.
  Step 1.2. Set $n = 1$.
Step 2. Iterative update.
  Step 2.1. Randomly generate an initial state $S^n_0 \in \mathcal{S}_0$.
  Step 2.2. Generate a sample path ω^n.
Step 3. The *forward* pass. Set t = 0.
  Step 3.1. For S^n_t, solve for $x^*_t = \arg\min_{x_t \in \mathcal{X}_t}\{\overline{V}^{n-1}_t(S^{x,n}_t)\}$.
  Step 3.2. Observe the state transition and record the visited post-decision states $S^{x,n}_t$.

$$S^{x,n}_t = S^{M,x}(S^n_t, x^*_t) = S^n_t + \Delta(x^*_t),$$
$$S^n_{t+1} = S^{M,W}(S^{x,n}_t, W_{t+1}(\omega^n)) = S^{x,n}_t + \epsilon(\omega^n_{t+1}),$$

  where $\Delta(x^*_t)$ is the expected growth, and $\epsilon(\omega^n_{t+1})$ is the random deviation from $\Delta(x^*_t)$.
  Step 3.3. If $F_{t+1} \geqslant 18mm$,
    mark day $t+1$ as the hCG day, i.e., $T = t+1$, and go to **Step 4**;
  else
    increment t and go to **Step 3.1**.
Step 4. The *backward* pass. For $t = T-1, \ldots, 0$, do
  Step 4.1. Retrieve the new observed value for the visited state S^x_t.
    Note that we decompose the state value to be the sum of $\hat{v}^{e,n}_t(P^e_t)$ and $\hat{v}^{o,n}_t(P^o_t)$.
  Step 4.2. Perturb P^e_t in $S^{x,n}_t$ by δ^e to become $\widetilde{P}^e_t = P^e_t + \delta^e$, simulate until the hCG day to obtain the perturbed cost on the hCG day, $\tilde{v}^{e,n}_{t,j}\left(\widetilde{P}^e_t\right)$.
    Update the visited segment j of the E_2 value function according to (12).
  Step 4.3. Similarly, perturb P^o_t in $S^{x,n}_t$ by δ^o to become $\widetilde{P}^o_t = P^o_t + \delta^o$, simulate until the hCG day to obtain the perturbed cost on the hCG day, $\tilde{v}^{o,n}_{t,j}\left(\widetilde{P}^o_t\right)$.
    Update the visited segment j of the ovary value function according to (13).
  Step 4.4. Smooth the slopes with stepsize as in (14) and (15).
  Step 4.5. Maintain the convexity or concavity in the value functions using the SPAR algorithm as in Powell (2007), and update $\bar{\psi}^{e,n}_{t,j}(P^e_t)$ and $\bar{\psi}^{o,n}_{t,j}(P^o_t)$.
Step 5. Let $n = n+1$. If $n \leqslant N$, go to **Step 2**.
Step 6. Return the piecewise linear (PWL) functions.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

4.4. An indexed piecewise linear function

In Section 4.3, we approximate the value of being in a state as a separable additive function with regard to the projected E_2 level and ovary diameter where we assume the values of these two dimensions are independent. Next, we try to explore the impact of their correlation using an indexed piecewise linear function.

We partition the ovary diameters into $\mathcal{K}$ intervals, $\mathcal{I}_k, k = 1, \ldots, \mathcal{K}$. We then use a family of value functions with respect to the projected E_2 level and index them by the ovary interval, i.e., $\overline{V}^{e,x,n}_{t,k}(P^{e,x}_t | O_t \in \mathcal{I}_k, k = 1, \ldots, \mathcal{K})$. That is, there is a different value function for the projected E_2 level $\overline{V}^{e,x,n}_{t,k}(P^{e,x}_t)$ associated with each different projected ovary interval k. By doing so, we somehow introduce the correlation between the E_2 level and ovary diameter.

However, we now have more functions to estimate and each indexed PWL function has less chance to be visited and updated, especially in early iterations. We may therefore suffer from a slower convergence. To overcome this, we adopt the hierarchical design as in Chapter 7 of Powell (2007). We consider the separable PWL function as the *aggregate* level and the indexed PWL functions as the *disaggregate* level. By combining both, we have

$$\overline{V}^{e,n}_t(P^e_t) = \left(1 - w^n_{t,k}\right)\overline{V}^{e,n}_{t,separable}(P^e_t) + w^n_{t,k}\overline{V}^{e,n}_{t,k}(P^e_t | O_t \in \mathcal{I}_k, k = 1, \ldots, \mathcal{K}), \quad (19)$$

where the weights $w^n_{t,k}$ are dynamically updated by considering the total variation (variance and bias of the transient data during the algorithm iterations, Chapter 7 of Powell, 2007). In this study, we use $\mathcal{K} = 6$.

5. Numerical experiments

In this section, we undertake numerical experiments to study the performance of the proposed ADP algorithms. We first study the convergence behavior of the three ADP algorithms and evaluate their resulting policies. We then study the impact of discretization levels in the lookup-table algorithm and stepsize rules in the separable PWL approximation algorithm.

Throughout the experiments, we use MDP with discretization level 14 as the benchmark. When generating the trivariate truncated Normal variables, we use the *normal-to-anything* (NORTA) procedure (Biller and Nelson, 2003, 2005; Cario and Nelson, 1996, 1998).

All the algorithms are programmed in C++ under Microsoft Visual Studio 2005 and all the experiments are conducted on a personal PC with Intel Core 2 Duo 3800 at 2.40 GHz CPU and 4.0 GB memory.

5.1. Comparisons of algorithms

We compare the three ADP algorithms in terms of their convergence behaviors, closeness to the MDP benchmark, and policy evaluation as the distributions of E_2 levels and ovary diameters on the hCG day.

5.1.1. Convergence behaviors

In Fig. 7, we plot the evaluated objective values of the three ADP methods in every 100 iterations up to 10000 iterations. Each evaluated objective value is the average simulated value of the same set of $M = 50$ initial states as in Section 3 and each initial state with 1000 independently generated sample paths.

We observe that both separable and indexed PWL algorithms stabilize at around 5000 iteration, while the lookup-table ADP improves very slowly and remains far away from the MDP benchmark until 10000 iterations. Fig. 8 plots the evaluated objective values of the lookup-table ADP up to 200 million iterations. The objective value converges slowly to the MDP benchmark.

We report the solution time in the second column of Table 3. All three ADP algorithms complete 10000 iterations in a few seconds. The lookup-table ADP takes the least time (however with a poor accuracy up to 10000 iterations), as it only needs to update the value of the visited state on a particular day t, whereas the two PWL approximations spend extra time to maintain concavity/convexity of the value function in addition to update the visited segment. With a family of E_2 value functions to update and maintain concavity/convexity, the indexed PWL ADP takes more time than the separable PWL ADP.

5.1.2. Policy evaluation

To evaluate the policies obtained at iteration 10000 of the three ADP algorithms, we use the average evaluated values of the same set of $M = 50$ initial states as in Section 3, each with 10000 independently generated sample paths. We provide the average cost on the hCG day and the standard deviation in the last two columns in Table 3.

Until 10000 iterations, there is still a 35.7% gap on the average cost between the lookup-table ADP and the MDP benchmark. The gap reduces to 4.2% at 200 million iterations, but it takes 12.62 h. On the other hand, the separable and indexed PWL ADP methods are 1.0% and 2.3% off the MDP benchmark with solution time in seconds (Notably, MDP takes about 41.2 h). The two PWL ADP algorithms significantly outperform the lookup-table ADP in terms of both solution accuracy and time. The indexed PWL ADP converges a little slower than the separable PWL ADP and demands more computational time in the value function updates. The results indicate that separable PWL value function is a good approximation of the value function in this dosage problem.

To further evaluate the policies obtained by the ADP algorithms, we present the distributions of E_2 levels and ovary diameters on the hCG day in Figs. 9 and 10 and their percentages relative to the target range in Table 4.

We observe that the two PWL ADP algorithms obtain policies that generate very similar distributions of the E_2 levels and ovary diameters on the hCG day as the policy of the MDP benchmark does, with the separable PWL ADP algorithm being closer to the MDP benchmark. The lookup-table ADP, with inadequate computational efforts by 10000 iterations, results in relatively lower E_2 levels and smaller ovary diameters, a sign of under-stimulation. Note that the ovaries of PCOS patients (the patient class under study in this paper) tend to be large. While under-stimulation may result in more ovaries in the target range, the E_2 levels tend to fall below the target range, which is undesirable as E_2 level is a more useful indicator of the stimulation level. This observation is consistent with the clinical literature and the observations in He et al. (2010).

The policy evaluation results indicate that the PWL ADP algorithms obtain policies with very close effect as the MDP benchmark, but with significantly less solution time.

5.2. Discretization levels in the lookup-table ADP

In the lookup-table ADP method, we need to discretize the continuous state space and the state transitions, which introduces discretization errors. The lookup-table ADP converges to optimality in theory if each state is visited infinitely often. However, we concern more on its finite time performance in practice. In Section 3, we have studied discretization levels in MDP when the discretized problems are solved exactly. Next, we experiment on the impact of discretization level in the lookup-table ADP within a fixed number of iterations.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

We experiment with the lookup-table ADP at discretization levels 1–14 as in Section 3. We plot the evaluated objective values in every 200000 iterations up to 20 million iterations in Fig. 11a. For presentation clarity, we only provide a subset of the convergence plots. We evaluate the policy obtained at the end of 20 million iterations and provide the evaluated objective values in Fig. 11b. As in previous experiments, the evaluation is done with simulation from the same set of $M = 50$ initial states as in Section 3. For each initial state, we use 1000 sample paths in the convergence plot in Fig. 11a and 10000 sample paths in the evaluation of the policy obtained at the end of 20 million iterations in Fig. 11a. In both Fig. 11a and b, the "MDP" line refers to the "MDP benchmark."

We observe that, with too coarse discretization levels (levels 1–7, represented by level 6), the objective values converge quickly but produce poor approximations. On the other hand, discretization levels that are too fine (levels 13 and 14, represented by level 14) result in very slow convergence due to the large state spaces and therefore slow updates. Their plots have not stabilized by 20 million iterations (note that level 14 converges in about 200 million iterations as in Fig. 8). With intermediate discretization levels (levels 8–12, represented by levels 10 and 12), the method appears to converge at reasonable rates without stalling at a poor approximation.

Our results highlight the tradeoff between discretization level and solution time. In practice, it may not be easy to decide on a proper discretization level. An alternative is to use a hierarchical approach (Chapter 7 of Powell, 2007). In this problem, however, the piecewise linear (PWL) function approximation methods offer a good solution with the utilization of the problem properties.

5.3. Stepsize rule

Next, we compare the harmonic and BAKF stepsize rules using the separable PWL ADP algorithm. Harmonic stepsize rule is non-adaptive/deterministic, while the BAKF is a stochastic stepsize rule that adapts to the state values collected as the algorithm proceeds. In Fig. 12, we plot the evaluated objective values up to 10000 iterations, under the BAKF stepsize rule and the harmonic stepsize rule with different parameter $a (a = 1, 10, 100)$ as in (9).

When a is small, the harmonic stepsize rule approximately degenerates to the $1/n$ rule, which generally does not perform well in practice and can create "false convergence," as the stepsize decreases to zero too quickly. This is illustrated with the cases when $a = 1$ and $a = 10$. A larger a (e.g., $a = 100$) prevents the stepsize from decreasing too fast in early iterations and can help to produce a good stream of convergence in a finite number of iterations. However, too large "a" will generate a long stream of large stepsizes (close to 1) and lead to over-sensitivity to new observations, which therefore results in large variance of the estimated statistics. In general, a proper value of the parameter "a" depends on the rate of convergence, which is hard to predict a priori.

The BAKF stepsize rule, however, is adaptive to the bias and variance of the transient stream of estimated state values as the algorithm proceeds and is therefore able to respond to different rates of convergence without having to tune parameters (Section 6.5.3 in Powell, 2007 and George and Powell, 2006). The adaptive nature of the BAKF stepsize rule to transient data streams makes it preferable in ADP.

6. Conclusions

As the first step in the *in vitro fertilization-embryo transfer* (IVF-ET) therapy, the *controlled ovarian hyperstimulation* (COH) cycle helps infertile women to produce multiple oocytes through gonadotropin administration to increase pregnancy probability. To avoid the iatrogenic complication *ovarian hyperstimulation syndrome* (OHSS), clinicians need to closely monitor the patients' physiological responses to gonadotropin administration and adjust dosages dynamically to avoid over-stimulation. However, the patients' responses are typically stochastic and unknown prior to the dosage administration. The stochastic and dynamic nature of the treatment cycle makes it difficult for the clinicians to be objective and consistent in their clinical decision making. The failure in treatment, though, can be life threatening in the case of severe OHSS as a result of over-stimulation, or extremely painful both physically and psychologically to the patients with no pregnancy as a result of under-stimulation.

With the availability of clinical data and experiences accumulated in years as well as the knowledge base in relevant clinical literature, one should be able to apply statistical tools to describe the physiological responses to gonadotropin administration of each patient sensitivity class, based on which the stochastic dynamic model and ADP methods proposed in this paper can assist clinicians to make more informed and consistent dosage decisions in the COH treatments, by avoiding subjective decision making or "the influence of the most recent cases." The significant solution time reduction of the proposed piecewise linear (PWL) approximation algorithms and the associated solution accuracy prove the success of the ADP algorithms to overcome the well-known curses of dimensionality in Markov decision processes (MDP). Furthermore, the flexibility of the ADP framework in modeling and solution algorithm development enables adaptation of clinical successes via integration of successful clinical policies or decision rules.

In recent years, both engineering and medical societies propose to improve health care delivery quality via the partnership between engineering and health care (Reid et al., 2005). This research provides a pioneer example of applying the cutting-edge operations research tool, i.e., approximate dynamic programming, in assisting evidence-based clinical practice and data-driven clinical decision making in the COH treatments. Schmid (2012) describes a recent application of approximate dynamic programming in the dynamic ambulance dispatching and relocation problem.

References

Aboulghar, M., 2003. Prediction of ovarian hyperstimulation syndrome: estradiol level has an important role in the prediciton of ohss. Human Reproduction 18 (6), 1140–1141.

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

Balasch, J., Fábregues, F., Creus, M., Puerto, B., Peñarrubia, J., Vanrell, J.A., 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), 652–656.

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

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

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

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

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

ESHRE, 2008. The European IVF monitoring programme (EIM), for the european society of human reproduction and embryology (eshre). Assisted reproductive technology in europe 2004. Results generate from european registers by ESHRE. Human Reproduction 23 (4), 756–771.

George, A., Powell, W.B., 2006. Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming. Machine Learning 65, 167–198.

He, M., Zhao, L., Powell, W.B., 2010. Optimal control of dosage decisions in controlled ovarian hyperstimulation. Annals of Operations Research 178, 223–245.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049

Heijnen, E.M.E.W., Macklon, N.S., Fauser, B.C.J.M., 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), 1936–1938.

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

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

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

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

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

Reid, P.P., Compton, W.D., Grossman, J.H., Fanjiang, G., 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.

Schmid, V., 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research 219, 611–621.

Seifer, D.B., Collins, R.L., 2002. Office-Based Infertility Practice. Springer.

Sutton, R., Barto, A., 1998. Reinforcement Learning. The MIT Press, Cambridge, MA.

Tarlatzis, B., 2002. Ovulation Induction. Elsevier.

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

Wikland, M., Bergh, C., Borg, K., Hillensjö, Howles, C.M., Knutsson, A., Nilsson, L., Wood, M., 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), 1677–1681.

Please cite this article in press as: He, M., et al. Approximate dynamic programming algorithms for optimal dosage decisions in controlled ovarian hyperstimulation. European Journal of Operational Research (2012), http://dx.doi.org/10.1016/j.ejor.2012.03.049