The Optimal Composition of Influenza Vaccines Subject to Random Production Yields

Soo-Haeng Cho[1]

UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095-1481
scho@anderson.ucla.edu

December 30, 2007

Abstract

The Vaccine and Related Biologic Products Advisory Committee meets at least once a year to decide the composition of the influenza vaccine in the U.S. Past evidence suggests that the Committee could use a more systematic approach to incorporate observed information and to quantify the risks associated with different options. There are two key tradeoffs involved in this decision. One, if the Committee decides to retain the current vaccine composition instead of updating to a new one, there is lower uncertainty in production yields, but the current vaccine could be less effective if a new virus strain spreads. Two, if the Committee decides early with less information, then manufacturers have more production time but the reduced information increases the risk of choosing a wrong strain. We derive an optimal dynamic policy for this decision. Due to the greater uncertainty in production yields of new vaccines, the optimal thresholds are neither symmetric between retaining and updating the composition nor monotonic over time. We apply our model to past decisions using parameter values estimated from a historical case. Our analysis shows that the dynamic optimal policy can significantly improve social welfare.

Key words: Dynamic Programming, Healthcare, Information, Product Development, Production Yield

[1] The author appreciates helpful suggestions from Kevin McCardle throughout this research. The author gratefully acknowledges useful conversations with Charles Corbett, Sarang Deo, and the CDC personnel including Duane Kilgus and Jack Nemecek regarding the vaccine demand and pricing issues. An earlier version was awarded the 2nd prize from the MSOM Student Paper Competition as well as the finalist award from the Decision Analysis Society Student Paper Competition in 2007.

1. Introduction

According to the World Health Organization (WHO), the burden of influenza in the U.S. is estimated to be 25-50 million cases per year, leading to 150,000 hospitalizations and 30,000-40,000 deaths (WHO 2007); this results in annual economic cost of $3-18 billion (WHO 2005). In light of the proven efficacy and safety of vaccines, the WHO recommends that all individuals should ideally have the opportunity to be vaccinated against influenza. The current influenza vaccine consists of three virus strains: one from each of subtypes A/H1N1 and A/H3N2, and type B. Influenza viruses constantly mutate. Unfortunately, vaccine immunity for one virus strain does not protect fully against subsequent variants. As a consequence, vaccine composition must be reviewed and, if necessary, changed every year to match circulating virus strains. To this end, the WHO has established a global surveillance network to monitor prevailing strains. In February, it recommends the annual vaccine composition for the northern hemisphere for the flu season which begins the following October.

In the U.S., the Vaccine and Related Biologic Products Advisory Committee (hereinafter, Committee) convenes soon after the WHO's announcement in order to make a final recommendation to the Food and Drug Administration (FDA) regarding the composition of vaccines. For each of the virus types, the Committee chooses among the three options: retain the current vaccine strain, change to a new strain, or defer a recommendation to the next meeting. Table 1 shows the decisions for each virus type during the past 6 years. In making this decision, the Committee analyzes both global and local surveillance data to answer the following four questions: (i) "are new influenza viruses present?", (ii) "are new strains spreading?", (iii) "do current vaccines induce antibodies against the new viruses?", and (iv) "are strains suitable for manufacturing?". If information is not sufficient to conclude a better strain, a final recommendation is deferred to collect more information, thereby reducing the risk of choosing a wrong strain. However, the delay of the recommendation may cause a supply shortage due to the short vaccination period, limited production capacity, and random production yields. We elaborate each of these points next.

Table 1: Vaccine Composition Decisions from 2002 to 2007.

	First Decision			*Final Decision*	
	Retain	Change	Defer	Retain	Change
A/H1N1	5	1	0	5	1
A/H3N2	2	3	1	3	3
B	3	1	2	3	3

The vaccination period is concentrated in October and November because this period provides a good balance between becoming immunized before the onset of the flu season and not being so early that immunity from a vaccine wanes [Centers for Disease Control and Prevention (CDC) 2007]. For example, in

2000 and 2001 a significant proportion of vaccine supply was distributed much later, resulting in substantial vaccine surpluses (O'Mara et al. 2003). Because unused vaccines have no salvage value, the timely delivery of vaccines is critical in this market.

The production facility is set and cannot be expanded during a flu season because its expansion takes 4 to 5 years (Matthews 2006). As of October 12, 2007, six manufacturers are projecting that as many as 132 million doses of vaccines will be available in the U.S. during the 2007-8 season (CDC 2007). Once vaccine composition is determined by the Committee, the manufacturers operate under a tight timeline for producing, testing, packaging, and distributing vaccines. Because of the long production and regulatory review processes, the manufacturers have to decide their target production quantities very early. Any problem encountered during production may cause a shortage or delay, and in fact, such problems have impacted the supply in 4 of the past 6 years (CDC 2007).

The production yields of strains are variable and unknown owing to its biological characteristic (Matthews 2006). Moreover, yield uncertainty is increased significantly when a vaccine strain is changed. The magnitude of this challenge is illustrated in the following quote from an industry representative (Committee 2003):

> "certainly the best way to ensure this predictability of supply is not to recommend any [strain] changes, ... a second best way is to minimize the number of strain changes. Each new strain can yield anywhere from 50 to 120 percent of the average strain."

Thus, even if a new virus strain is predominant, a change is made only when the benefit from improved efficacy outweighs the risk associated with making the change in production. For instance, although new A/Fujian-like virus strains were widely spreading during the 2002-3 season, the Committee did not select that strain because it was not suitable for large-scale manufacturing.

As described, the vaccine composition decision is quite complex because of the multiple decision criteria involved as well as the myriads of uncertainties in virus activities and vaccine production. The past evidence suggests that the Committee could use a more systematic approach to incorporate observed information and to quantify the risks and costs associated with the various options. For example, in 2004, the Committee deferred a final recommendation of the type B strain even though 52 (91%) virus isolates of a new candidate strain (belonging to B/Yamagata lineage) had been collected and 5 (9%) of the current strain (belonging to B/Victoria lineage). Because of the predominance of one strain, several members raised the concerns such as (Committee 2004):

> "if you just put a 95% confident interval on that [91%] crudely, it's compatible with something between 83% and 100%, as the true underlying rate would lie within that interval. So I think that with data we have it would take a lot to overturn it. … if we assume that we got 10 more B isolates, and

that they were all not Yamagata, you still would have observed 63% [52/67 = 78% is correct] of them, and you'd still be more than 50 …"

To aid this complex decision process, we develop a dynamic decision model as a two-stage game. Because the strain selection decision is largely independent among the three virus types, we focus on the decision of a single type to simplify our analysis. In the first stage, the Committee chooses among the three options described earlier: retain, change or defer. There are two key tradeoffs involved in this decision. The first tradeoff is between selecting the current strain and selecting a new strain: if the Committee decides to retain the current vaccine strain, there is lower uncertainty in production yields, but the vaccine could be less effective if a new virus strain spreads. The second tradeoff is between selecting a strain early and selecting a strain late: if the Committee decides early with less information, then manufacturers have more production time in the second stage, but the reduced information increases the risk of choosing a wrong strain. In the second stage, the manufacturers engage in quantity competition to produce vaccines containing the chosen strain. They independently decide their target production quantities within the capacity constrained by the available production time. The actual quantities produced are determined by random production yields. A market-clearing price is set according to a linear demand function.

Our analysis shows that there is an optimal threshold policy for when to retain the current strain, change to a new strain, or defer. The greater yield uncertainty of a new strain leads to a smaller quantity of vaccines in equilibrium, so it is optimal for the Committee to collect more information before selecting a new strain. This drives the structural property that the optimal thresholds are neither symmetric between the two strains nor monotonic over time. We conduct numerical analyses to investigate both the magnitude and direction of the impact of various factors on the optimal policy such as: the cross effectiveness of vaccines, production yield uncertainty, entry of new manufacturers, capacity expansion of incumbent manufacturers, the development of a new technology, and the severity and progress of virus activities. Lastly, we apply our model to the Committee's 2004 decisions described above. Our analysis shows that the practice of the Committee meeting only once or twice on pre-determined dates could result in a welfare loss of $29 million compared to the optimal dynamic policy. This suggests that the timing of meetings needs to be flexible to cope with constantly changing situations. We also find that policies which enhance the consumers' awareness of indirect costs of infection can improve social welfare substantially and that policy interventions in vaccine price are necessary to achieve the CDC's target size of vaccinated population.

The remainder of the paper is organized as follows. In section 2, we review related literature. In section 3, we present our model and discuss how our model captures the Committee's current decision criteria. In section 4, we derive an optimal policy and discuss its structural property. In section 5, we investigate the effect of changes in various parameters on the optimal policy. In section 6, we apply our model to

the Committee's decisions in 2004. In section 7, we summarize the insights from our analysis, and discuss limitations and extensions. We conclude this paper in section 8.

2. Related Literature

This work draws on and contributes to two distinct streams of literature. First, we extend the literature on new product introduction decisions under uncertainty to a competitive setting where both demand and supply are uncertain. Second, we show in the influenza vaccine context how uncertainty in production affects the optimal timing of the vaccine composition decision and the consequent social welfare.

The influenza vaccine composition decision can be considered as the product upgrade decision made annually. Several papers (e.g., Cohen et al. 1996, Bayus 1997, Morgan et al. 2001) address the trade-offs between the timing of new-product introductions, product quality, and product development costs. While they do not account for uncertainties involved in this decision-making, Souza et al. (2004) and Özer and Uncu (2006) consider the product introduction and subsequent production decisions under demand uncertainty. In Souza et al. (2004), both product introduction and production decisions are made simultaneously in a period, and a firm maximizes the total expected discounted profit generated from multiple product generations over an infinite horizon. Özer and Uncu (2006) is particularly germane to our work. They model the decision of introducing a single new-generation product for a component supplier in a high-tech industry. They develop a two-stage model: in the first stage, a firm decides when to stop process design and enter the market, and in the second stage, a firm makes production decisions to satisfy uncertain demand. Production yields are imperfect but predictable, and depend on the time of market entry. The literature in the adoption of a new technology also shares some features with the above stream of research. In McCardle (1985) and Ulu and Smith (2007), for instance, a firm decides whether or not to adopt a new technology with unknown benefit. It may defer the decision and gather more information in order to reduce the risk of adopting an unprofitable technology.

In this paper we also address the product upgrade and production decisions sequentially. The main difference is that a decision-maker in our model faces uncertainties in both demand and supply in making a product-upgrade decision. In the first stage of our model, a decision-maker collects information about uncertain demand – i.e., the Committee collects information about virus prevalence, which in turn determines the efficacy of vaccines and the demand for vaccines. The decision in the first stage affects production decisions in the second stage in two ways: its timing determines production capacity until a short selling season, and its composition – whether or not an upgrade is made – determines the degree of uncertainty in production yields. In making production decisions, firms engage in quantity competition. We contribute to the literature by characterizing the impact of random production yields on the product upgrade decision in a competitive setting. While our model is most appropriate to influenza vaccine, the

analysis in this paper can be extended to other types of products or technologies with unknown demand and supply in the industries such as pharmaceutical, semiconductor, and agriculture. Examples include cost-reducing technologies in a semiconductor industry having different production yields, and new design alternatives in a high-tech industry having different degrees of manufacturability.

Influenza vaccine problems have been studied extensively. The main research streams have been the development of epidemic models to predict the spread of infectious diseases (see reviews in Bailey 1975, Anderson and May 1991, and Earn et al. 2002) and the evaluation of policy interventions to control them (e.g., Brito et al. 1991, Philipson 2003). Recent vaccine shortages have led operations researchers to study the role of uncertain production yields in supply. Federgruen and Yang (2007) analyze a model for a firm or public organization to decide a set of suppliers and the amount of orders from each under uncertain demand and supply. Chick et al. (2006) show that if a central government is the sole purchaser and a single manufacturer supplies all vaccines at a fixed price, a cost-sharing contract guarantees a sufficient supply of vaccines. As the authors admit, most epidemic models (including theirs) assume that the government can precisely specify the number of individuals to be vaccinated, ignoring the fact that individuals make their own immunization decisions. Corbett and Deo (2006) employ a competitive model of multiple manufacturers and show that yield uncertainty can contribute to a high degree of industry concentration. This paper builds upon and contributes to this line of research by analyzing the vaccine composition decision, a primary source of unpredictability in supply, in the realistic settings where multiple manufacturers engage in quantity competition, outputs are subject to random production yields, vaccine price increases as a smaller quantity of vaccines is supplied, and individuals make their own vaccination decisions.

The importance of selecting *right* vaccine strains has received attention only recently. Wu et al. (2005) show that the current generic vaccine performs well relative to a hypothetical vaccine customized to an individual based on his/her antigenic history. Kornish and Keeney (2006) address the same decision problem of the Committee as the current paper. They employ a finite-horizon optimal stopping model and show that there exists an optimal policy between selecting a strain now and deferring in order to collect more information. We extend Kornish and Keeney (2006) by accounting for the following important characterstics: (i) vaccine is effective but imperfect, and induces antibodies that cross-react to different strains; (ii) vaccine demand is nondecreasing in vaccine efficacy that depends a match between vaccine strains and circulating virus strains; (iii) vaccine production is subject to random production yields; (iv) uncertainty in production is greater when producing new vaccines; (v) multiple competitive manufacturers make production decisions. Modeling these factors explicitly allows us to accommodate the Committee's current decision criteria and to make precise arguments for their impacts on the optimal policy. By applying the model to a recent past decision, we compare the performance of various policies and draw insights on the current practice.

3. Model

3.1 Stage 1: vaccine composition

We employ a finite-horizon discrete-time model with $t \in \{1, 2, \ldots, T\}$, where '$t$ = 1' represents the time when the Committee is first convened (usually, February 16~20), and '$t = T$' when the production campaign should end in order for vaccines to be distributed (usually before October 1). Reasonable values for time intervals are between one week and four weeks.

Prevalence of virus strains Consistent with the current decision-making process, the Committee selects between the current vaccine strain, $s1$, and the most prevalent new strain, $s2$. The prevalence of $s1$ is measured by θ ($\in [0, 1]$), the proportion of $s1$ cases among all cases caused by either strain during the upcoming flu season. The true value of θ is unknown until the season ends. The Committee's beliefs in period t about θ are contained in the random variable $\tilde{\theta}_t$. We assume that the initial distribution of $\tilde{\theta}_1$ can be represented by a beta distribution with parameters (α, β). The beta distribution is an appropriate model for the proportion θ because it not only models a random variable between zero and one, but can also express a variety of prior attitudes by adjusting the parameters. The Committee could develop this initial distribution by using multi-variant epidemic models such as Castillo-Chavez et al. (1989) or Andreasen et al. (1997). Based on the Committee meeting transcripts, however, it seems that the Committee does not use any forecast model (Committee 2003-2006) perhaps because of the difficulty in estimating parameters in those models – in particular, the parameters that capture seasonal trends (Earn et al. 2002).

In each period the Committee observes new information about virus spread. At time τ, for t = 1, 2, …, τ-1, let N_{t+1} ($\in N \cup \{0\}$) denote the amount of information observed during the period (t, t+1], and for t = τ, τ+1, …, T-1, let N_{t+1} denote the estimate of the amount of information forthcoming in the future period (t, t+1]. The past data indicate that a peak of influenza activity is usually reached before the first meeting of the Committee, and then the number of new cases decreases gradually (see Figure 1). One could estimate N_{t+1} based on this pattern or use epidemic models (e.g., Bailey 1975, Anderson and May 1991). We assume that the number of $s1$-favorable pieces of information forthcoming in the future period (t, t+1], M_{t+1}, follows a binomial distribution with parameters (N_{t+1}, θ). We show later in this section that this assumption implies that *on average* there are no anticipated changes in the forecast.

The Committee updates its prior distribution in a Bayesian fashion with new information. Since the beta prior and Bernoulli sampling form a conjugate pair, the resulting posterior distribution is also a beta. If we let $\hat{\theta}$ denote the current mean of $\tilde{\theta}_t$ in period t, then the pre-posterior mean in period t+1 is

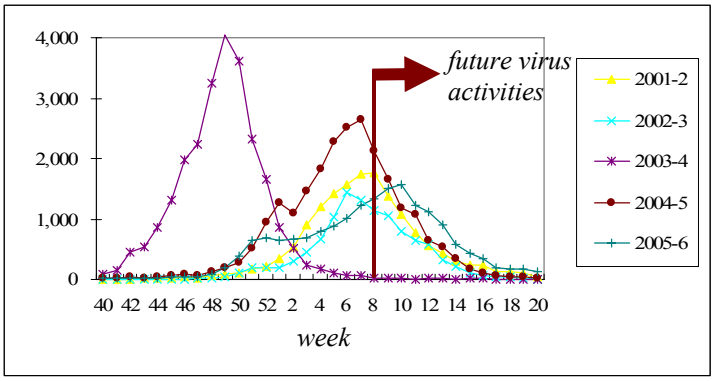

Figure 1: Weekly Number of Virus Specimens Tested Positive by the WHO and National Respiratory and Enteric Virus Surveillance System collaborating laboratories in the U.S. (CDC 2007).

$$\Theta_{t+1}(\hat{\theta}) = \frac{(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)\hat{\theta} + M_{t+1}}{\alpha + \beta + \sum_{\tau=2}^{t+1} N_\tau}. \tag{1}$$

The unconditional probability distribution of M_{t+1} is a beta-binomial as the prior distribution of $\tilde{\theta}_t$ is a beta with parameters $\{(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)\hat{\theta}, (\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)(1-\hat{\theta})\}$. Then the probability distribution of $\Theta_{t+1}(\hat{\theta})$ is:

$$\Pr\left\{\Theta_{t+1}(\hat{\theta}) = \frac{(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)\hat{\theta} + m}{\alpha + \beta + \sum_{\tau=2}^{t+1} N_\tau}\right\} = \binom{N_{t+1}}{m} \frac{B\left\{(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)\hat{\theta} + m,\ (\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)(1-\hat{\theta}) + N_{t+1} - m\right\}}{B\left\{(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)\hat{\theta},\ (\alpha + \beta + \sum_{\tau=2}^{t} N_\tau)(1-\hat{\theta})\right\}} \tag{2}$$

where $B(a, b)$ is a beta function with parameters (a, b). The prior expected value and variance of $\Theta_{t+1}(\hat{\theta})$ are obtained as follows (e.g., Morris 1968):

$$E[\Theta_{t+1}(\hat{\theta})] = \hat{\theta} \tag{3}$$

$$Var[\Theta_{t+1}(\hat{\theta})] = \frac{N_{t+1}}{(\alpha + \beta + \sum_{\tau=2}^{t+1} N_\tau)} Var[\tilde{\theta}_t] = \frac{N_{t+1}\hat{\theta}(1-\hat{\theta})}{(\alpha + \beta + \sum_{\tau=2}^{t+1} N_\tau)(\alpha + \beta + \sum_{\tau=2}^{t} N_\tau + 1)}. \tag{4}$$

Equation (3) implies that the current forecast $\hat{\theta}$ is built upon all available information so that it will remain the same *on average* in the next period. A similar assumption is made in Kornish and Keeney (2006). Equation (4) shows that the prior variance of the posterior mean increases as less information has been collected until the current period and more information is forthcoming in the next period. Note that $Var[\tilde{\theta}_t]$ is decreasing in t with positive N_t for any fixed $\hat{\theta}$. This implies that more information tightens the posterior distribution and reduces the uncertainty about the prevalence.

Vaccine efficacy The efficacy of a vaccine depends on a match between the vaccine strain and circulating virus strain (e.g., Nichol 2003). Let e_{jk} denote the efficacy of a vaccine containing strain j against virus strain k ($j, k \in \{1, 2\}$). Note that e_{jk} with $j \neq k$ reflects how well vaccine j makes antibodies that cross-react to virus k, and that e_{12} need not equal e_{21}. While a good estimate of e_{12} can be obtained by testing the current vaccine, e_{21} is less known because a vaccine containing a new strain does not exist yet. However, the degree of cross-reaction can be identified serologically (Levine 1992). Nichol (2003) reports that vaccine efficacy does not depend much on age and health conditions although it is slightly lower for the elderly. For simplicity, we assume that e_{jk} is known and constant for all individuals.

We define the overall efficacy of vaccine j, e_j, by $e_j(\theta) \equiv e_{j1}\theta + e_{j2}(1-\theta)$. Then, the expected efficacy in period t is: $E\, e_j(\tilde{\theta}_t) = e_j(\hat{\theta}) = (e_{j1} - e_{j2})\hat{\theta} + e_{j2}$. Since a vaccine is effective but imperfect, and its efficacy is greater when a vaccine strain matches a circulating virus, we assume that e_{jk} is between 0 and 1 with $\min(e_{11}, e_{22}) > \max(e_{12}, e_{21})$. Then, $e_j(\hat{\theta})$ is positive for all $\hat{\theta}$, $e_1(\hat{\theta})$ is increasing in $\hat{\theta}$, and $e_2(\hat{\theta})$ is decreasing in $\hat{\theta}$ with $e_1(0) < e_2(0)$, and $e_1(1) > e_2(1)$.

Dynamic programming formulation In each period t, the Committee faces three alternatives: select $s1$, select $s2$, or defer. The objective is to maximize expected social welfare. Let $W_{t,j}(\hat{\theta})$ denote the expected social welfare (of which the definition is presented in the next section) if the Committee selects strain j at time t when the current estimate of θ is $\hat{\theta}$. Let $V_t(\hat{\theta})$ be the expected social welfare following an optimal policy from t on with estimate $\hat{\theta}$. Define $W_t(\hat{\theta}) \equiv \max\{W_{t,1}(\hat{\theta}), W_{t,2}(\hat{\theta})\}$. Then $V_t(\hat{\theta})$ satisfies the following dynamic programming recursion for $t = 1, 2, \ldots, T\text{-}1$:

$$V_t(\hat{\theta}) = \max\{W_t(\hat{\theta}),\ EV_{t+1}(\Theta_{t+1}(\hat{\theta}))\} \tag{5}$$

The first term in the maximization is the expected social welfare from selecting a strain now, and the second term is that from deferral. Note that there is no *direct* cost of searching for new information in (5) because surveillance activities continue regardless of the Committee's decision. The terminal condition is set as $V_T(\hat{\theta}) = 0$ as no vaccines would be available in time for consumption.

3.2 Stage 2: vaccine production and consumers' vaccination

Vaccine supply Suppose the Committee has selected strain j at time t. We assume that there are n identical manufacturers indexed by $l \in \{1, 2, \ldots, n\}$, where n is fixed as the number of manufacturers who are licensed to produce vaccines is well known in advance. Each manufacturer's production capacity is proportional to the available production time, which is determined in turn by when the Committee selects a

strain. Each manufacturer processes embryonated eggs into vaccines at a constant rate r, so its production capacity is $r(T-t)$. Within this capacity, manufacturer l chooses the number of embryonated eggs $\bar{q}_{t,j,l}$ to maximize its expected profit. The marginal cost of processing each egg is c_1.

Following Chick et al. (2006) and Corbett and Deo (2006), the quantity of vaccines produced is represented as the product of the number of eggs put into production and a random production yield. Let $\tilde{y}_{j,l}$ denote the random yield of manufacturer l in producing vaccine j. The total quantity of vaccines $\tilde{q}_{t,j}$ is then $\sum_{l=1}^{n} \tilde{y}_{j,l}\bar{q}_{t,j,l}$. We assume that $\tilde{y}_{j,l}$ is independently and identically distributed for all manufacturers with mean μ_j, standard deviation σ_j, and coefficient of variation δ_j (= σ_j/μ_j). We assume the high-growth reassortants and potency test reagents, which are required to achieve moderate yields in producing trivalent vaccines, are available to manufacturers for both strains in a timely manner, hence $\mu_1 = \mu_2$ (= μ). As the yield is more predictable when producing vaccines containing the current strain, we assume $\sigma_1 < \sigma_2$, hence $\delta_1 < \delta_2$.

Note that without a high-growth reassortant, the selection of a new strain is prohibitive because the use of a wild-type strain will result in a significantly low production yield. For example, A/Fujian-like strain was not selected in 2003 for this reason. It is easy to see that if the average yields of the two strains are different, a strain with a higher average yield is, *ceteris paribus*, more favorable.

Vaccine demand We consider two cost components that affect an individual's decision regarding vaccination. The first, $\tilde{c}_2$, is the expected cost of infection incurred by an individual if he/she is not vaccinated. It reflects the likelihood of getting infected and the consequent costs, including both direct costs of health care and indirect costs due to work loss and death. We assume for analytical tractability that $\tilde{c}_2$ is uniformly distributed between 0 and $2c_2$ in the population. The second, c_3, is the indirect cost incurred by an individual if he/she gets vaccinated. It includes pain, lost work time for vaccination, and the cost of potential side effects. We assume that c_3 is the same for everyone.

One dose of vaccine is recommended for people older than 6 months except for children younger than 9 years old who receive vaccine for the first time (Nichol 2003). So, we assume that each individual consumes at most one dose of vaccine every year. There are two types of consumers: informed and uninformed. Informed consumers have beliefs about vaccine efficacy that match those of the Committee. This type includes those who decide upon vaccination following the recommendation of their physicians. Uninformed consumers, on the other hand, decide vaccination based on their individually perceived efficacy $\bar{e}$. We assume that $\bar{e}$ is the same for every uninformed consumer as in Brito et al. (1991), and that the composition of a vaccine does not affect $\bar{e}$. A randomly-chosen individual is informed with probability x, which is exogenously given in our model.

Let $v^{(i)}$ ($v^{(u)}$) denote the value of vaccination of an informed (uninformed) consumer. Then, $v^{(i)} = e_j(\hat{\theta})\,\tilde{c}_2 - c_3$ and $v^{(u)} = \bar{e}\,\tilde{c}_2 - c_3$. If the price of a vaccine is p, then individuals with values above p get vaccinated. Thus, the demand D in the population of size S is given by

$$D(p) = xS\left(1 - \frac{p + c_3}{2c_2 e_j(\hat{\theta})}\right) + (1-x)S\left(1 - \frac{p + c_3}{2c_2\bar{e}}\right). \tag{6}$$

When the total quantity of vaccines $\tilde{q}_{t,j}(\hat{\theta})$ is introduced to a market, a market-clearing price p is determined according to the inverse demand function as in Corbett and Deo (2006):

$$p(\tilde{q}_{t,j}(\hat{\theta})) = \left\{\frac{2c_2\bar{e}e_j(\hat{\theta})}{x\bar{e} + (1-x)e_j(\hat{\theta})} - c_3\right\} - \frac{2c_2\bar{e}e_j(\hat{\theta})}{S\{x\bar{e} + (1-x)e_j(\hat{\theta})\}}\tilde{q}_{t,j}(\hat{\theta}). \tag{7}$$

Note that while manufacturers choose their production quantities to maximize their expected profits based on their anticipated demand *ex ante*, the market may not clear *ex post* because of inefficient market-clearing mechanism as well as random production yields. Essentially, we assume a linear demand function in the form of $p = a - bq$ where parameters a and b are constructed using a simple model of consumers' immunization decisions. It is clear from the following excerpt from GAO (2001) that vaccine price increases as a smaller quantity of vaccines is supplied during the short demand window:

> "During the shortage period, providers who wanted to purchase vaccine often faced rapidly escalating prices from distributors with an available supply. For example, orders placed by physicians in our sample during the peak vaccination months of October and November cost an average of $7 per dose, compared with less than $3 per dose for orders that had been placed before the end of June 2000. State health officials and providers who had placed orders early often waited for delivery of their lower-priced vaccine from manufacturers and distributors until late November or December."

Social welfare Social welfare is defined as the sum of the utilities over both vaccinated and unvaccinated populations less the production cost (i.e., the sum of consumers' and producers' surpluses). Although uninformed consumers decide whether or not to get vaccinated based on their perceived efficacy $\bar{e}$, their utilities depend on the actual efficacy $e_j(\theta)$. Let $\tilde{c}^{(i)}$ ($\tilde{c}^{(u)}$) denote the expected cost of infection of the informed (uninformed) consumers who are indifferent between getting vaccinated and not. Then the expected social welfare when the Committee selects strain j at time t is given by

$$\begin{aligned} W_{t,j}(\hat{\theta}) = & -E\left[xS\int_{\tilde{c}^{(i)}}^{2c_2}\{(1-e_j(\hat{\theta}))z + c_3\}\frac{dz}{2c_2}\right] - E\left[xS\int_0^{\tilde{c}^{(i)}} z\,\frac{dz}{2c_2}\right] \\ & - E\left[(1-x)S\int_{\tilde{c}^{(u)}}^{2c_2}\{(1-e_j(\hat{\theta}))z + c_3\}\frac{dz}{2c_2}\right] - E\left[(1-x)S\int_0^{\tilde{c}^{(u)}} z\frac{dz}{2c_2}\right] - c_1\sum_{l=1}^{n}\bar{q}_{t,j,l}(\hat{\theta}) + c_2 S. \end{aligned} \tag{8}$$

The first four terms in (8) represent the expected utilities of different groups: informed vaccinated, informed unvaccinated, uninformed vaccinated, and uninformed unvaccinated, respectively. The fifth term represents the variable cost of production. We add the mean cost of infection incurred by unvaccinated population, $c_2 S$, in order to have the terminal condition '$V_T(\hat{\theta}) = 0$' without vaccines. We do not include in (8) the fixed costs of surveillance, R&D and production as they would occur whichever strain is chosen. The welfare function in (8) does not capture the positive externalities of vaccines explicitly (see detailed discussion in section 7). For simplicity, we do not consider distributors' margins.

3.3 Discussion of the current decision criteria

When a new virus strain antigenically distinguishable from the current vaccine strain is spreading, the Committee employs three decision criteria in selecting a strain; each of these criteria plays a significant role in our model. The first criterion, "are new strains spreading?", is captured in our model by θ, the proportion of influenza cases due to the current vaccine strain. The second criterion, "do current vaccines induce antibodies against the new viruses?", is captured by e_{12}, the cross-efficacy of the current vaccine. The third criterion, "are strains suitable for manufacturing?" is captured by $\tilde{y}_{j,l}$, the random production yield. The Committee defers a recommendation when information is not sufficient to conclude a "better" strain and additional significant information is likely to be forthcoming. This is modeled by the updating of θ. A sequence of N_t determines the rate at which the estimation of θ becomes more accurate.

4. Analysis

We solve the model using backward induction in three steps. In the first step, we solve the second stage game among vaccine manufacturers, given that the Committee has selected strain *j* at time *t* when its estimate of θ is $\hat{\theta}$. In the second step, we find the optimal strain to select at each time *t*. In the final step, we derive the subgame-perfect optimal policy of when the Committee should stop gathering information and select the optimal strain found in the previous step.

4.1 Stage 2: vaccine supply in equilibrium

In the second stage, each manufacturer chooses the number of embryonated eggs within its processing capacity in order to maximize its expected profit.

Proposition 1 *The expected total quantity of vaccines in equilibrium when the Committee has selected strain j at time t with the estimate* $\hat{\theta}$ *is given by*

$$E[\,\tilde{q}_{t,j}(\hat{\theta})\,] = \begin{cases} \dfrac{n\,S}{n+1+2\,\delta_j^2}\left[1 - \dfrac{c_3 + c_1\mu^{-1}}{2c_2}\left\{\dfrac{x}{e_j(\hat{\theta})} + \dfrac{1-x}{\bar{e}}\right\}\right] \;(\equiv q_j(\hat{\theta})) & \text{if } t \le t_j(\hat{\theta}) \\ n\,r\mu\,(T-t) & \text{if } t > t_j(\hat{\theta}) \end{cases} \tag{9}$$

$$\text{where } \; t_j(\hat{\theta}) \equiv T - \frac{S}{r\mu(n+1+2\delta_j^2)}\left[1 - \frac{c_3 + c_1\mu^{-1}}{2c_2}\left\{\frac{x}{e_j(\hat{\theta})} + \frac{1-x}{\bar{e}}\right\}\right]. \tag{10}$$

Moreover, $q_j(\hat{\theta})$ *is increasing in* $e_j(\hat{\theta})$ *for x > 0 and in* $\bar{e}$ *for x < 1, but decreasing in* δ_j.

(All proofs are provided in Appendix A. We assume $q_j(\hat{\theta}) > 0$ for all j, x, and $\hat{\theta}$.)

Proposition 1 shows that vaccine supply decreases by deferring strain selection only when available production time is less than T-$t_j(\hat{\theta})$. It seems intuitive that manufacturers would not produce more than anticipated demand even if they could. With greater efficacy, unconstrained manufacturers would produce more in response to greater demand. In contrast, unconstrained manufacturers would produce less when there is greater uncertainty in production (this result is consistent with Corbett and Deo 2006). Therefore, a smaller quantity of vaccines of the new strain *s*2 will be supplied in equilibrium than of the existing strain *s*1 (provided both vaccines are expected to have the same efficacy).

We first consider a case where all consumers are uninformed, *i.e.* $x = 0$, and then a case where informed consumers exist, *i.e.* $x > 0$.

4.2 Stage 1: optimal policy when $x = 0$

When all consumers are uninformed, randomness in vaccine efficacy does not influence consumers' decision about vaccination. So, the demand and supply of vaccines do not depend on how likely vaccine strains will match circulating virus strains. Substituting $x = 0$ into (9) and (10), we see that q_j is independent of $\hat{\theta}$ and that q_1 is *always* greater than q_2 due to lower yield uncertainty of *s*1. Accordingly, thresholds t_1 and t_2 are independent of $\hat{\theta}$ with $t_1 < t_2$.

The social welfare, however, still depends on random efficacy because the utilities of consumers depend on the actual efficacy $e_j(\theta)$, not on the perceived efficacy $\bar{e}$. The discrepancy between these efficacies might result in more vaccinated population than the Committee would expect at the social optimum. However, considering that the vaccination target population size of 218 million (CDC as of July 17, 2006) is much higher than the current vaccine supply of about 100 million doses, it is reasonable to assume that vaccine supply in equilibrium is less than a socially-optimal quantity in expectation. Under this assumption, we establish the following results:

Proposition 2 *When all consumers are uninformed,*

(a) $W_{t,j}(\hat{\theta})$ *is linearly increasing in* $e_j(\hat{\theta})$ $\forall\, t, j,$ *and* $\hat{\theta}$*;*

(b) $W_{t,j}(\hat{\theta})$ *is increasing in* $E[\tilde{q}_{t,j}]$ *and decreasing in* δ_j $\forall\, t, j,$ *and* $\hat{\theta}$.

The results in Proposition 2 are consistent with our general notion that more vaccines with greater efficacy are more beneficial to society. What is interesting is that the expected social welfare decreases as yield uncertainty increases whether or not manufacturers are capacity-constrained. When vaccines containing either strain are expected to have the same efficacy, therefore, it is optimal for the Committee to select the currently used strain *s*1. When manufacturers are capacity-constrained, social welfare as well as vaccine supply decreases as less time is available for production.

Next, we find the optimal strain $j_t^*(\hat{\theta})$ to select at time *t* with $\hat{\theta}$ if deferral is not an option. We define θ^a and θ^b as: (i) θ^a solves $e_1(\theta^a) = e_2(\theta^a)$, and (ii) θ^b solves $W_{t,1}(\theta^b) = W_{t,2}(\theta^b)$ when $t \le t_1$. Then;

Theorem 1 *When all consumers are uninformed, there exists a threshold number* θ_t^S *in each period t such that if the Committee selects a strain at t, then it is optimal to select s1 when* $\hat{\theta} \ge \theta_t^S$*, and to select s2 when* $\hat{\theta} \le \theta_t^S$*. Furthermore,* θ_t^S *equals* θ^b *for* $t \le t_1$*, and then increases to* θ^a *at T.*

Figure 2 (a) illustrates the results of Theorem 1. The four regions are divided by the thresholds t_1, t_2, and θ_t^S. Each decision region has a distinct pair $(j_t^*(\hat{\theta}),\ E[\tilde{q}_{t,j_t^*(\hat{\theta})}])$: $(1,\ q_1)$ in R_1^U, $(1,\ nr\mu(T-t))$ in R_1^C, $(2,\ q_2)$ in R_2^U, and $(2,\ nr\mu(T-t))$ in R_2^C. The decision regions are not symmetric between the two strains. Two major characteristics that distinguish the two strains drive this asymmetry: (i) biologic characteristics affecting vaccine efficacy, and (ii) operational characteristics of vaccine production. If vaccines containing either strain are equally effective, *i.e.* $e_{11} = e_{22}$ and $e_{12} = e_{21}$, then $\theta^a = 0.5$. If the production yields of both strains are equally uncertain, *i.e.* $\delta_1 = \delta_2$, then $t_1 = t_2$ and $\theta_t^S = \theta^a = \theta^b$ for all *t*. This asymmetry leads to an optimal policy that differs significantly between the two strains.

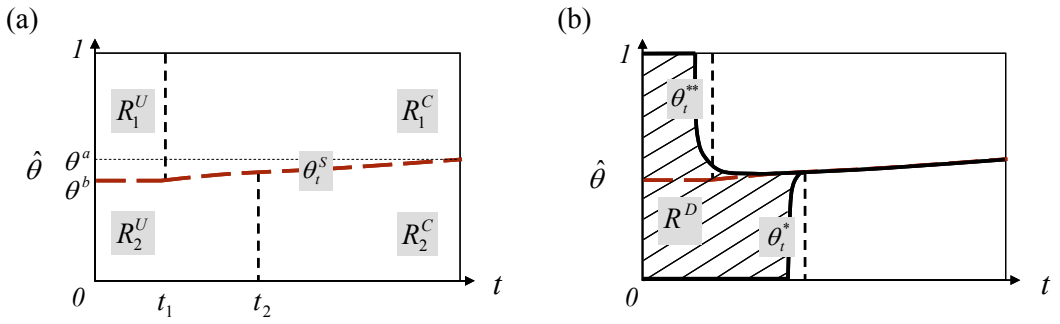

Figure 2: Decision Regions when *x* = 0: (a) Optimal Strain, and (b) Optimal Policy.

In each period $t = 2,\ldots, T\text{-}1$, the state space of $\hat{\theta}$ is $\{\lambda/(\alpha+\beta+\sum_{\tau=2}^{t} N_\tau) \mid \lambda = \alpha, \alpha+1, \ldots, \alpha+\sum_{\tau=2}^{t} N_\tau\}$; thus, the dynamic program has a finite time horizon with a finite state space. Given any value of the parameters, therefore, it is possible to find the optimal solution of (5) by computing the value function $V_t(\hat{\theta})$ for each possible value of $\hat{\theta}$ from $t = T\text{-}1$ to 1 backwards. There is, however, much more that can be said about the form of an optimal policy.

Since there is no direct cost of gathering additional information, the expected cost of deferring the selection of a strain is zero when vaccine supply is unaffected by deferral. When the manufacturers are not expected to be capacity-constrained in every possible state of the next period, it is always beneficial to defer the selection because more information about future virus activities reduces the risk of choosing a wrong strain. Formally, for any given t and $\hat{\theta}$, if every possible value of $\Theta_{t+1}(\hat{\theta})$ belongs to the region $(R_1^U \cup R_2^U)$, then deferral is optimal. This gives sufficient grounds for deferral to be optimal, and as such provides an inner envelope of the optimal deferral region.

Even if vaccine supply is expected to decline because of capacity constraints, deferral can still be optimal when there is a high risk that the currently optimal strain will differ from the one in the future. In order that such a risk exists, it is necessary that additional information should be able to alter the optimal strain in the future. Formally, if the minimum possible value of the state at $T\text{-}1$, starting from $\hat{\theta}$ at t in R_1^C, is greater than θ_{T-1}^S, then it is optimal to select $s1$ immediately. Similarly, if the maximum possible value of the state at any τ ($> t$), starting from $\hat{\theta}$ at t in R_2^C, is less than θ_τ^S, then it is optimal to select $s2$ immediately. This gives necessary grounds for the immediate selection to be optimal, and as such provides an outer envelope of the optimal deferral region.

The above analysis of the optimal deferral region suggests the existence of a pair $(\theta_t^*, \theta_t^{**})$ of threshold numbers in each period t such that if $\theta_t^* < \hat{\theta} < \theta_t^{**}$, it is optimal to continue collecting information about virus activities. However, it suggests little about the shape of the decision regions mainly due to the asymmetry manifested in Figure 2 (a). The next theorem shows that a threshold policy is indeed optimal, and describes how the thresholds change over time.

Theorem 2 *When all consumers are uninformed, there exists a pair of numbers θ_t^* and θ_t^{**} with $\theta_t^* \le \theta_t^S \le \theta_t^{**}$ in each period t such that if $\theta_t^* < \hat{\theta} < \theta_t^{**}$, it is optimal to defer; if $\hat{\theta} \ge \theta_t^{**}$, it is optimal select s1, and otherwise, it is optimal to select s2. If $\Theta_{t+1}(\hat{\theta})$ is riskier than $\Theta_{t+2}(\hat{\theta})$ in the sense of second-order stochastic dominance for $t > t_2 - 1$ given any $\hat{\theta}$, then θ_t^* is nondecreasing in t but θ_t^{**} is not monotone in t; both eventually converge to θ_t^S.*

Figure 2 (b) illustrates a typical optimal policy when $x = 0$, in which R^D represents the optimal deferral region. It shows that the decision regions for optimal strains shown in Figure 2 (a) have a significant effect on the optimal policy. This is because the cost of deferral is captured by t_1 and t_2, and the benefit of deferral is captured by θ_t^S. Since demand is inelastic to random efficacy, the optimal thresholds do not depend much on $\hat{\theta}$. The optimal policy is rather simple: defer until manufacturers are expected to be capacity-constrained in the next period and then select the optimal strain (unless $\hat{\theta}$ is very close to θ_t^S). The optimal deferral region disappears soon after t passes t_2 because production loss outweighs the benefit from little additional information during a late flu season.

The lower threshold θ_t^* is nondecreasing in t in a normal situation where N_t does not increase in t during a late flu season. However, the upper threshold θ_t^{**} is not generally monotone in t. This seems counterintuitive because the more information a decision-maker has, the more confident he/she should be about his estimate (McCardle 1985). One would expect, for instance, that if it is optimal to select $s1$ with the estimate $\hat{\theta}$ at t, the Committee would also find the selection of $s1$ to be optimal at t+1 with more information. This non-monotonicity occurs in our model because the threshold between selecting $s1$ and $s2$, θ_t^S, is time-dependent, whereas a break-even point between the adoption and rejection of a single technology, as in McCardle (1985), is time-invariant.

Figure 2 (b) reveals that the optimal policy is not symmetric between the two strains. The optimal selection date is later when the optimal strain is a new one. This suggests that the Committee should collect more information before choosing the new strain $s2$ than the currently used strain $s1$. The greater yield uncertainty in large-scale manufacturing leads to a smaller quantity of vaccines containing $s2$ in equilibrium, so the Committee has to be more careful before selecting the new strain. If both strains have the same biological and operational characteristics, the optimal policy is symmetric about $\hat{\theta} = 0.5$, and monotone over time as in Kornish and Keeney (2006).

4.3 Stage 1: optimal policy when $x > 0$

When informed consumers exist, the expectation about random vaccine efficacy affects the demand. While consumers with high or low valuation about vaccines may not change their decisions, informed consumers with moderate valuation will get vaccinated only when they expect high efficacy. We first analyze a case where $x = 1$ and then a case where $x \in (0, 1)$.

The following proposition summarizes the properties of the expected social welfare $W_{t,j}(\hat{\theta})$.

Proposition 3 *When all consumers are informed,*

(a) $W_{t,j}(\hat{\theta})$ *is increasing in* $e_j(\hat{\theta})$ *convexly when* $t < t_j(\hat{\theta})$, *and linearly when* $t \geq t_j(\hat{\theta})$ $\forall j$ *and* $\hat{\theta}$;

(b) Proposition 2 (b) and Theorem 1 are also valid.

When manufacturers are capacity-constrained, better efficacy does not lead to more vaccinated population; thus, the expected social welfare increases *linearly* in the expected efficacy. On the other hand, when manufacturers are not capacity-constrained, better efficacy not only improves the benefit of vaccination per consumer but also increase the size of vaccinated population, so the expected social welfare increases *convexly* in the expected efficacy. Figure 3 (a) illustrates decision regions for the optimal strain. Substituting $x = 1$ into (10), we see that $t_1(\hat{\theta})$ is decreasing and convex in $\hat{\theta}$, and $t_2(\hat{\theta})$ is increasing and concave in $\hat{\theta}$. This is because as $\hat{\theta}$ approaches θ_t^S, the expected efficacy becomes lower, resulting in less demand of informed consumers.

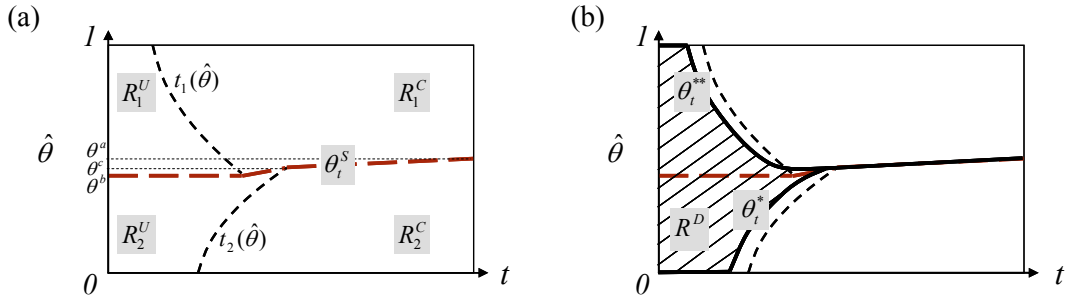

Figure 3: Decision Regions when $x = 1$: (a) Optimal Strain, and (b) Optimal Policy.

Unlike the case where $x = 0$, $W_t(\hat{\theta})$ is quasi-convex but not convex for $t \in (\min\{t_1(1), t_2(0)\}, t_2(\theta^c))$ where θ^c is the point at which t_2 crosses θ_t^S as shown in Figure 3 (a). Due to the complexity involved in computing the expected value of quasi-convex functions, unfortunately, we can make only an intuitive argument about the optimal thresholds during this period. In our computational study, we have found no exception of Theorem 2 in this case.

Figure 3 (b) illustrates a typical optimal policy when $x = 1$. As in the earlier case, the thresholds of the optimal strain, t_1, t_2 and θ_t^S, have a significant effect on the optimal policy, and the optimal thresholds are also asymmetric between the two strains and nonmonotonic over time. Because demand is elastic to $\hat{\theta}$, the optimal thresholds vary much with respect to $\hat{\theta}$. This clearly illustrates the benefit of consumers' awareness of random vaccine efficacy: it permits the Committee to employ a *flexible* policy. When one strain is predominant, *i.e.* $\hat{\theta}$ is close to 0 or 1, the Committee can optimally select that strain without any delay, which allows manufacturers more time to produce a large quantity of vaccines. On the other hand, when the Committee is highly uncertain of which strain is going to prevail, *i.e.* $\hat{\theta}$ is close to 0.5, the

Committee should optimally defer selection and collect more information, which reduces the risk of choosing a less effective strain.

When both informed and uninformed consumers are present, *i.e.* $x \in (0, 1)$, one would expect that an optimal policy has thresholds between those when $x = 0$ and 1. The intuition for this result is as follows: the more consumers are informed, the more elastic demand is to the likelihood of a match between the vaccine and virus strains, and the more flexible policy the Committee can implement. Our computational study bears this out. Figure 4 shows a typical example of optimal policies for different values of x.

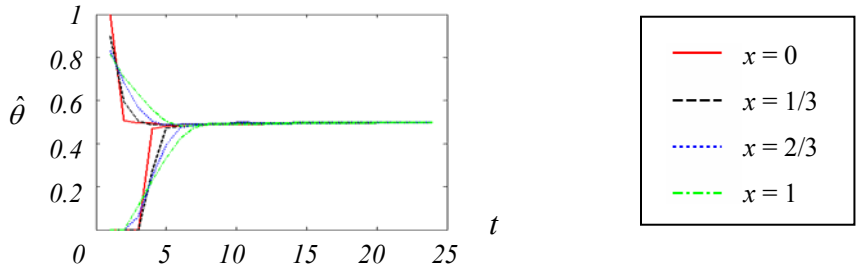

Figure 4: An Example of Optimal Policies for Different Values of x when $\bar{e} = e_1(0.8) = e_2(0.2)$.

(Note: Threshold curves are not smooth because they are computed in discrete state and time spaces. The same explanation applies to Figures 5 and 6.)

5. Comparative Statics

In this section we investigate the effect of changes in various parameters on the optimal policy. In order to show both the magnitude and direction of the effects, we present computational results for the case of x = 0.5, using realistic parameter values in Table 2. (We provide justifications for these values in the next section as we use the same values in applying the model to the situation in 2004.) The same arguments apply to all other cases of x. We examine parameter changes in the following order: (i) parameters that affect either $W_{t,1}$ or $W_{t,2}$ exclusively (e_{12} and δ_2), (ii) parameters that affect both $W_{t,1}$ and $W_{t,2}$ (r, n and c_2), and (iii) a sequence of N_t that affects V_t but not W_t.

Table 2: Base Parameter Values Used in Numerical Analyses.

Parameter	Value	Parameter	Value
$e_{11} = e_{22}$, $e_{12} = e_{21}$, $\bar{e}$	0.75, 0.35, 0.67	S (millions)	300
$\mu_1 = \mu_2$, σ_1, σ_2	0.9, 0.1, 0.4	n	2
c_1, c_2, c_3 (\$)	3, 18, 7	T (weeks)	25
(α, β); $N_1, N_2, ..., N_T$	(1,1);57,20,18,16,14,12,10,9,8,7,6,5,4,3,2,1,…,1	r (millions/week)	2.7778

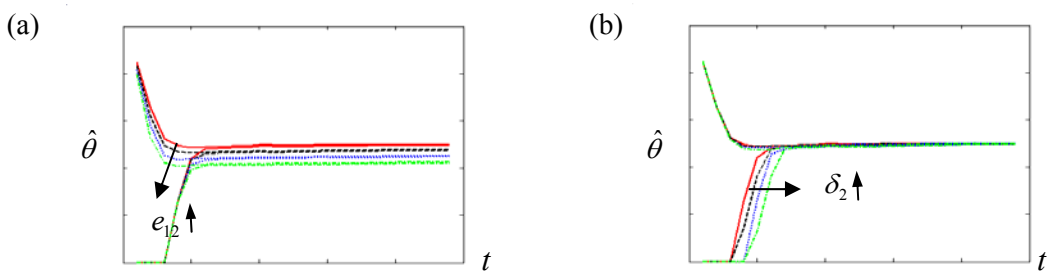

Figure 5: Examples of Optimal Policies: (a) e_{12} = 0.350 (base), 0.385 (+10%), 0.430 (+20%) and 0.465 (+30%); and (b) δ_2 = 0.444 (base), 0.489 (+10%), 0.533 (+20%) and 0.578 (+30%).

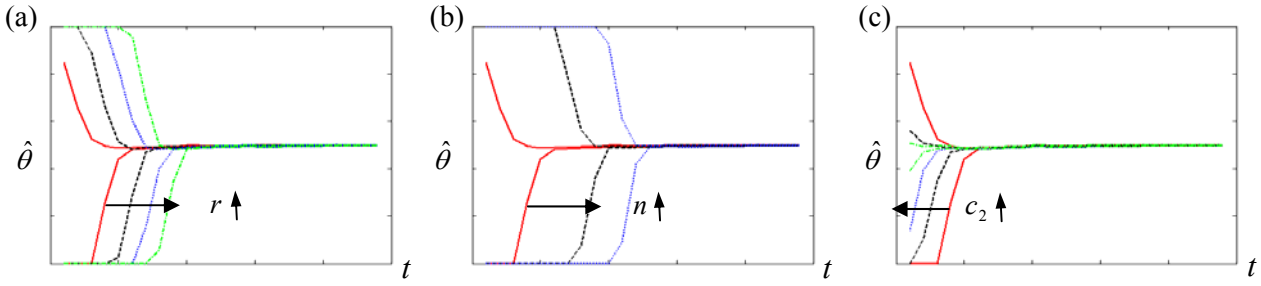

Figure 6: Examples of Optimal Policies: (a) r = 2.778 (base), 3.056 (+10%), 3.333 (+20%) and 3.611 (+30%); (b) n = 2, 3 and 4; and (c) c_2 = 18 (base), 19.8 (+10%), 21.6 (+20%) and 23.4 (+30%).

The cross-efficacy of the current vaccine, e_{12}, is the principal element the Committee considers when a new strain is spreading. Because of the greater uncertainty in production yields of new vaccines, the Committee tends to retain the current vaccine strain as long as it induces antibodies against the new viruses reasonably well; the greater cross-efficacy makes the expected vaccine efficacy less sensitive to the prevalence, $\hat{\theta}$. As e_{12} increases, therefore, we expect that the optimal decision region for selecting $s1$ expands, while other two regions contract. Figure 5 (a) shows that this intuition is true. The upper threshold remains fixed at $\hat{\theta}$ = 1, at which the expected efficacy does not depend on e_{12}, and then decreases more sharply in t as e_{12} increases.

The manufacturability of vaccines containing a new strain has been the primary concern of the Committee when a new strain is predominant but not well-protected by the current vaccine. Because the greater yield uncertainty leads to less supply in equilibrium, the Committee could defer longer before selecting a strain. Figure 5 (b) shows that as δ_2 increases, the optimal deferral region for the new strain s2 expands.

New technologies such as the use of tissue culture cells would accelerate the time for obtaining seed viruses. This would result in a shorter lead time and greater production rate r. Capacity expansion of incumbent manufacturers would also increase r. With an increase in r, the Committee could defer strain selection longer without risking a vaccine shortage as shown in Figure 6 (a).

As more manufacturers enter the market, the quantity of vaccines in equilibrium increases. Since $0 < \partial t_j / (\partial n / n) < \partial t_j / (\partial r / r)$ from (10), we expect that the optimal deferral region expands with additional (identical) manufacturers, but it does so less than with the same percent increase in *r*. This is illustrated in Figure 6 (b) where the increase of *n* from 2 to 3 (50% increase) expands the optimal deferral region to the extent equivalent to approximately 30% increase of *r* as shown in Figure 6 (a). Capacity expansion of incumbent manufacturers is more efficient than a new firm building up the same capacity.

Suppose the influenza attack rate in the upcoming season is expected to be higher than in a typical season. Then the expected cost of infection incurred by an unvaccinated individual increases. This would drive more demand for vaccines, so the Committee would be less likely to defer strain selection. Figure 6 (c) shows that the optimal deferral region contracts as c_2 increases.

Comparing the results shown in Figures 5 and 6, we see that the effects of changes in e_{12} and δ_2 are smaller than in *r*, *n* and c_2.

Lastly, we discuss the impact of the amount of new information in each period *t*, N_t, on the optimal policy. Consider two sequences $N_{t+1}^{(1)}$ and $N_{t+1}^{(2)}$ for $t = 1, \ldots, T\text{-}1$, and their corresponding pre-posterior means $\Theta_{t+1}^{(1)}(\hat{\theta})$ and $\Theta_{t+1}^{(2)}(\hat{\theta})$ given $\hat{\theta}$. Suppose $\Theta_{t+1}^{(1)}(\hat{\theta})$ is riskier than $\Theta_{t+1}^{(2)}(\hat{\theta})$ in the sense of second-order stochastic dominance for all *t* given any $\hat{\theta}$. Let $V_t^{(1)}$ and $V_t^{(2)}$ denote value functions with $N_{t+1}^{(1)}$ and $N_{t+1}^{(2)}$, respectively, when all other parameters are the same. When $x = 0$, $V_{t+1}(\hat{\theta})$ is convex in $\hat{\theta}$, hence $EV_{t+1}^{(1)}[\Theta_{t+1}^{(1)}(\hat{\theta})] > EV_{t+1}^{(2)}[\Theta_{t+1}^{(2)}(\hat{\theta})]$ for all *t*. This suggests that the greater uncertainty in prevalence makes deferral more attractive. When $x > 0$, however, $V_{t+1}(\hat{\theta})$ is no longer convex in $\hat{\theta}$, so the optimal deferral region is not necessarily larger with a riskier sequence.

6. Application

In this section, we apply our model to the Committee's 2004 decision regarding the type B strain discussed earlier. Recall that 57 virus isolates had been collected from October 1 in the previous year until the first meeting, of which 52 contained the new candidate strain while only 5 contained the current vaccine strain. While several members questioned the benefit of doing so, the Committee deferred a recommendation to the next meeting (scheduled 4 weeks later) for fear of a possible late outbreak and the small number of the collected isolates. Our objective in this section is not to argue whether or not the deferral was optimal in hindsight, but to illustrate the utility of our model and to compare the performance of various policies. We begin with calibrating our model. Table 2 summarizes parameter values in a base case. Some parameters do not have available references and in those instances our estimates are somewhat

rough. The effects of changes in key parameters are already discussed in the previous section. Sensitivity analyses for other parameters are also available upon request.

6.1 Model Calibration

Prevalence of virus strains We assume that the Committee had initial beliefs prior to the season that both strains would be equally likely to spread. This corresponds to (α, β) = (1, 1). (See section 6.3 for more explanation.) The epidemiology of two representative lineages of type B – B/Victoria and B/Yamagata – still remains poorly understood (Mizuta et al. 2004). While B/Yamagata lineage viruses predominated between 1990 and 2001, viruses from both lineages co-circulated since then. For this reason, it seems from the Committee meeting transcripts that the Committee selects a vaccine strain, hoping that the prevalence of the current season will match that of the next season. Thus, we assume that the Committee believes that each future virus isolate contains $s1$ with the probability θ.

We regard each virus isolate as one random sample. The number of the virus isolates collected before the first Committee meeting (N_1) is 57 (Committee 2004). We choose N_2 of 20 to reflect the concerns of the Committee about a possible late outbreak. Since a peak of virus activities is reached before the Committee's first meeting in a typical year (see Figure 1), we choose the remaining N_t so that it gradually decreases over time. Despite the Committee's concerns, it turned out that there was no late outbreak (CDC 2007).

Vaccine efficacy We assume a vaccine containing either strain (B/Yamagata or B/Victoria) would be equally effective. Based on several references, Nichol (2001) finds that the average vaccine efficacy for healthy working adults is 0.75 (or 0.35) in a year with a good (or poor) match between the vaccine and virus strains, and that the likelihood of a good match is 0.8. Since vaccine efficacy does not vary much among the population (Nichol 2003), we choose $e_{11} = e_{22} = 0.75$, $e_{12} = e_{21} = 0.35$, and $\bar{e} = 0.75 \times 0.8 + 0.35 \times 0.2 = 0.67$.

Vaccine supply Before the British regulatory authorities prohibited the sale of Chiron's vaccines, Aventis and Chiron produced 55 and 48 million doses, respectively (Offit 2005). Thus, we choose the number of manufacturers (n) of 2. The assumption of symmetric manufacturers also seems reasonable. Since manufacturers have two different production campaigns for 6 months (Gerdil 2003), we choose T = 25 (weeks), assuming 24 weeks of continuous production. The production rate (r) of 2.7778 (million/week) is chosen to have the industry capacity of 120 million doses. We choose the industry capacity based on the evidence that the total quantity produced during the entire season was 61 million doses (Corbett and Deo 2006).

The growth characteristic of the B/Victoria strain ($s1$) was "moderate", while that of the B/Yamagata strain ($s2$) was less known but considered "reasonable" (Committee 2004). So, the assumption of identi-

cal average yields between the two seems appropriate. Unfortunately, the mean value of the moderate yield is not directly available, so we estimate it based on the following quotes (Committee 2001, 2003):

> "we actually found that at the end of the day the yield is still significantly lower than the B/Yamanashi [having the moderate growth characteristic], it is about 65 to 70%, in fact"
>
> "one low-yielding strain could mean as much as 20 million doses fewer for the market"

The first quote indicates that the mean of the low yield is about 70%, and the second quote indicates that if we assume a total vaccine supply of 100 million doses, the mean of the low yield is about 20% lower than that of the moderate yield. Thus, we set μ to 0.9 in a base case.

The yield uncertainty is less observable. Chick et al. (2006) use 0.16~0.45 for the standard deviation of a random yield, and Corbett and Deo (2006) use 0~3 (0.64 in a base case) for the coefficient of variation. The estimates of Corbett and Deo (2006) seem a bit high, considering that the coefficient of variation of a uniform random variable between 0 and 1.8 is 0.58. Note that these papers do not distinguish the yield of the current vaccine strain from that of a new candidate strain. Since it is said that a yield is "predictable" if retaining the current vaccine composition (Committee 2004-6), we choose σ_1 of 0.1. We estimate σ_2 based on the following quote (Committee 2003):

> "each new strain can yield anywhere from 50 to 120 percent of the average strain"

Let $\tilde{y}_j$ denote the random variable representing the yield of strain *j*. If we assume $\tilde{y}_2$ follows a uniform distribution between $0.5\,\tilde{y}_1$ and $1.5\,\tilde{y}_1$, then the standard deviation of $\tilde{y}_2$ is 0.3. Since the above quote does not take account of the possibility of a complete production failure, which happened to Chiron in 2004, we choose σ_2 of 0.4 conservatively (see Figure 5 (b) for the effect of σ_2 on the optimal policy).

Cost components We use the same unit production cost (c_1) of \$3 as Corbett and Deo (2006). We set the expected cost of infection among the population (c_2) to \$18, using the estimates of Weycker et al. (2005) for different age and health-risk groups. Details about this choice are presented in Appendix B. Nichol (2001) estimates an indirect cost of vaccination of \$6.69 for healthy working adults. For simplicity, we choose the indirect cost of vaccination (c_3) of \$7.

6.2 Equilibrium Analysis

In this section, we examine the expected quantity of vaccines and the consequent social welfare when the current vaccine composition is retained or updated. We consider a case when $\bar{e} = e_1(0.8) = e_2(0.2)$, at which the proportion of informed consumers, *x*, does not affect the results. Figure 7 (a) shows that at $t = 1$, the manufacturers are not capacity-constrained and the expected quantity of vaccines is 113 million (or 101 million) if the current composition is retained (or updated). The manufacturers are capacity-

constrained from t = 3 (or 5) on if the current composition is retained (or updated). Figure 7 (b) shows that the difference in the expected social welfare between the two options is $100 million at t = 1 or 2, and then decreases to $30 million at t = 5. This implies that the degree of production yield uncertainty has a significant influence on the quantity of vaccines and social welfare in equilibrium. Consequently, the optimal timing of making a composition decision also depends on whether the current vaccine composition is retained or not. We have already presented the optimal policy of this case in Figure 4. It shows that the optimal timing of selecting the new strain $s2$ is 2~3 weeks later than that of retaining the current vaccine strain $s1$.

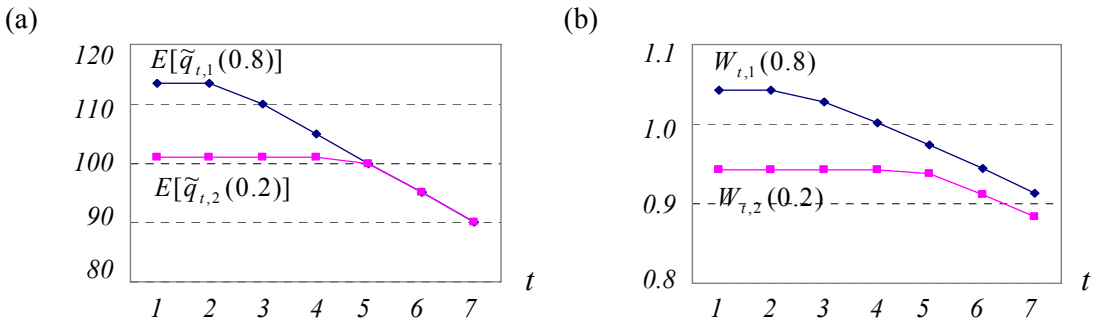

Figure 7: Impact of Production Yield Uncertainty on: (a) Expected Quantity of Vaccines (million doses); and (b) Expected Social Welfare (billion $) when $\bar{e} = e_1(0.8) = e_2(0.2)$.

6.3 Policy Evaluation

We first examine the likelihood of the current vaccine strain becoming predominant with new isolates forthcoming during the next 4 weeks. We consider prior distributions in the beginning of the season with (α, β) being equal to: (i) (1, 1), (ii) (32, 13), and (iii) (7, 3). Case (i) represents beliefs that both strains would be equally likely to spread. This seems reasonable because type B vaccine strains had been changed 13 times between 1980 and 2003 (Gerdil 2003, Committee 2002-3). Case (ii) represents beliefs based on the pre-season data: the CDC had collected 45 virus isolates between April and September in 2003, of which 32 contained the current vaccine strain and 13 contained the new strain. Since the pre-season data are less informative than the actual season data, the Committee might have had less confidence about its initial forecast. Case (iii) represents one instance of this scenario with a lower value of α + β. The Committee's prior is biased toward the current vaccine strain in the cases (ii) and (iii) as illustrated in Figure 8. At the first meeting, the Committee's beliefs are updated using 57 isolates, resulting in the prevalence estimates for the current vaccine strain of 0.10, 0.36 and 0.18 for the cases (i), (ii) and (iii), respectively. Table 3 shows that, irrespective of the amount of new information, the chances that the current strain would become predominant after 4 weeks are very low in all cases. The number of collected isolates and the possibility of a late outbreak should not have been reasons for deferral.

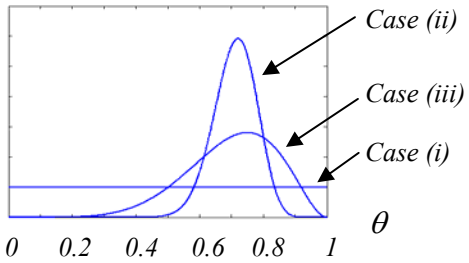

Figure 8: Prior Probability Distributions of the Prevalence for the Current Vaccine Strain.

Table 3: Probabilities of the Posterior Mean Being Greater Than or Equal to 0.5.

(α, β) \ $\sum_{\tau=2}^{5} N_\tau$	10	50	200	500
(1, 1)	0	9.2E-24	5.9E-14	1.8E-12
(32, 13)	0	7.3E-07	3.5E-04	1.1E-03
(7, 3)	0	4.1E-18	1.8E-10	3.2E-09

Next, we compare the past decisions to the optimal decisions from the model. We consider the following two policies: *Policy A* represents a dynamic policy of monitoring virus activities every week and making decisions accordingly; *Policy B* represents the current policy of holding the first meeting and then, if necessary, having another meeting after 4 weeks, and so on. We use the same basic parameter values in Table 2 for both policies except that $T = 25$ (weeks) and $r = 2.7778$ (million/week) in *Policy A*, while $T = 7$ (periods) and $r = 4 \times 2.7778$ (million/period) in *Policy B*. The optimal decisions of *Policy A* are illustrated in Figure 4. It shows that it was indeed optimal to defer at the first meeting ($t = 1$) with the Committee's prevalence estimate for the current strain of 0.09. This is mainly because the manufacturers were not expected to be capacity-constrained in the near future [confirmed during the meeting by an industry representative, who said, "this time, I don't see a particular problem with that. That's not always true …" (Committee 2004)]. However, our analysis shows that it was optimal in *Policy B* to select the new strain at the first meeting because the Committee does not make any use of new information until the second meeting after 4 weeks. The expected loss of social welfare due to deferral ranges from $6 million when $x = 0$ to $29 million when $x = 1$. In the second meeting, the Committee updated its estimate to 0.20 (Committee 2004). Since the further delay of production might cause a supply shortage, it was optimal in both policies to select the new strain.

Lastly, we evaluate the impacts of the demand-side and supply-side policies on social welfare. In Appendix B, we analyze vaccine coverage in the subgroups separated by age and health conditions. It shows that individuals do not internalize their indirect costs of infection in making their vaccination decisions.

Policies that enhance awareness of such costs will improve social welfare by increasing demand. We consider an increase of c_2 (from a base of \$18) to \$28 or \$41 when individuals fully internalize their indirect costs due to work loss or, additionally, future earnings lost due to death. A supply-side policy we consider is to increase production capacity. Figure 9 shows that expected social welfare increases by \$1.41 billion or \$3.13 billion with the base capacity when c_2 = \$28 or \$41, respectively. Additional capacity improves social welfare only when it can satisfy unmet demand, but it does so less than the demand-side policy. With a 20% increase of r from the base case, expected social welfare increases by \$0.11 billion or \$0.47 billion when c_2 = \$28 or \$41, respectively. Thus, the demand-side policy seems more effective than the supply-side policy. Finally, note that when c_2 = \$41, the equilibrium demand increases to only 144 million (as opposed to the CDC target size of 218 million) because of a subsequent price increase. If the price is maintained through government subsidies, our model predicts the demand of 211 million.

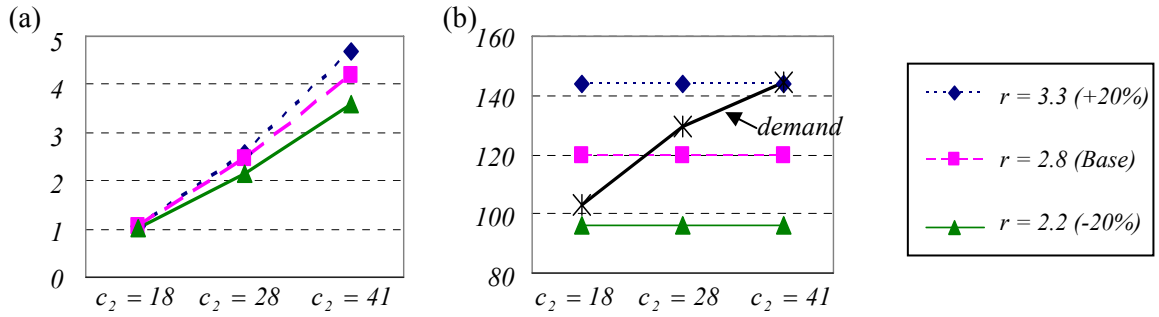

Figure 9: Impact of Demand- and Supply-side Policies on: (a) Expected Social Welfare (billion \$); and (b) Capacity and Demand (million doses) when $T = 25$, $\hat{\theta} = 0.1$, $t = 1$, and $x = 0.5$.

7. Discussion

In this paper we have derived an optimal policy for the influenza vaccine composition and subsequent production decisions. The greater yield uncertainty of a new strain leads to a smaller quantity of vaccines in equilibrium, so it is optimal for the Committee to collect more information before updating the vaccine composition. This drives the structural property that the optimal thresholds are neither symmetric between the two strains nor monotonic over time. As the demand for vaccines is more elastic to vaccine efficacy, the Committee can implement a more flexible policy depending on its forecast about a match between the vaccine and virus strains. This result justifies the tremendous efforts of the CDC to publicize surveillance information. However, more efforts are required to educate the public about how to incorporate surveillance information into their vaccination decisions. We also analyze the impact of various factors on the optimal policy such as: the cross effectiveness of vaccines, production yield uncertainty, entry of new manufacturers, capacity expansion of incumbent manufacturers, the development of a new technology,

and the severity and progress of virus activities. By incorporating these key components, our model allows the Committee to perform what-if analyses under various scenarios.

The key difference of our model-based approach from the current practice is that we treat the decision-making process as a *dynamic* process. Currently, the Committee meets only once or twice on predetermined dates. Our analysis has shown that the optimal policy varies depending on several factors including virus activities, efficacies, yields, and capacities that may differ every season. Thus, the timing of meetings needs to be flexible to cope with constantly changing situations. Especially, once deferral is decided, the Committee needs to update its beliefs about future virus prevalence more frequently. To this end, a consistent rule needs to be set up regarding how to distribute weights between the local and global surveillance data, and between the early and late season data. We also find that policies which enhance awareness of indirect costs of infection can improve social welfare substantially, but policy interventions in vaccine price are necessary to achieve the CDC target size of vaccinated population.

Our model is also capable of assessing the consequences of policy changes in vaccine composition. For instance, the possibility of adding one more strain of type B or subtype A/H5N1 (avian flu) has been discussed extensively. One of the key concerns with a quadravalent vaccine is whether the demand increase from improved efficacy is sufficient to sustain the additional capacity needed. By accounting for the effects of vaccine efficacy on demand and supply, our model can provide insight to this issue.

While our model captures key components in the vaccine composition decision, we make several simplifying assumptions to maintain tractability.

One, we focus on the decision of a single virus type as in Kornish and Keeney (2006). The strain selection decision is largely independent among the three types except that manufacturers utilize the same production facility in producing monovalent vaccines of each type. A model extension to the three types is left for future work.

Two, we assume that the Committee is risk-neutral and maximizes the expected social welfare. In actuality, the Committee might be risk-averse toward choosing a wrong strain but not necessarily toward a vaccine shortage, for the role of the Committee's composition decision is more transparent in the former than in the latter. Intuitively, incorporation of risk aversion into the model would expand the optimal deferral region. Alternatively, behavioral bias may explain the Committee's deferral decision in 2004. For instance, the Committee might have a tendency to defer too long by overestimating the quality of late information (see, e.g., Bearden et al. 2006).

Three, we use a simple aggregate demand model without taking account of unique characteristics of subgroups in the population. This modeling choice is made to focus on key issues in the vaccine composition decision, which is the decision of the most upstream player in the supply chain. Modeling heterogeneity becomes more important in analyzing decisions in the downstream of the supply chain such as the

optimal allocation of vaccines among subgroups (e.g., Hill and Longini 2003). Also, we assume a single vaccine price while pre-season orders may receive some discounts.

Four, our demand and welfare functions do not capture the positive externalities of vaccines *explicitly*. Brito et al. (1991) accounts for them by representing the expected cost of infection as the product of the cost of infection and the probability of getting infected which decreases as more people are vaccinated. We argue that these effects on the vaccine composition decision are second-order for the following three reasons: (i) the expected vaccine quantity in equilibrium resides in a narrow range (in the example in section 6, 101~113 million doses when $x = 0$) so that the proper choice of the parameter c_2 can *implicitly* capture the externalities to some degree, (ii) the effects of the externalities are common to both candidate strains, and (iii) as demand increases, the probability of getting infected decreases, which in turn creates a counteracting feedback in the demand (Philipson 2003); Jefferson et al. (2005) found no empirical evidence of the association between the vaccine coverage and the attack rate of influenza-like illness (with a correlation coefficient of 0.09).

Five, our model does not specify how the Committee *should* develop its beliefs about the prevalence of the upcoming season based on the dynamics of influenza evolution. The incorporation of a forecast model such as a multi-variant SIR model (e.g., Andreasen et al. 1997) into our decision-theoretic model would be ideal. To this end, the empirical tests of models of drift evolution must precede (Earn et al. 2002).

Six, while our model accounts for uncertainties in virus activities and production yields, we assume complete information for others such as N_t, e_{jk}, μ, and σ_j. Sensitivity analyses would mitigate this limitation.

8. Conclusions

In this paper we have developed and analyzed a dynamic model for the Committee's decision of influenza vaccine composition. The decision-making is quite complex due to the multiple decision criteria and uncertainties in several dimensions. The Committee expends a great deal of effort to collect all relevant information but, unfortunately, does not fully utilize that information in a systematic way. By incorporating the Committee's current decision criteria and quantifying the risks associated with different options, our model and analysis provide practical insights to support the decision-making of the Committee.

Most work in the literature of influenza vaccines pertain to the issues related to the downstream of the vaccine supply chain. These include the forecast and control of influenza through policy interventions in vaccinations. Only recently, operations researchers have addressed the issues in the upstream of the supply chain such as the composition and production of vaccines. However, they have treated these two issues separately. A main contribution of this paper is to characterize the impact of the vaccine composition

decision on the subsequent decisions in the supply chain and to provide novel insights that could not be obtained otherwise.

The influenza vaccine composition decision can be considered as the product upgrade decision made annually. While our model is most appropriate to influenza vaccine, the analysis in this paper can be extended to other industries such as pharmaceutical, semiconductor, and agriculture, in which random production yields play an important role in making the product upgrade and subsequent production decisions.

Appendix

A. Proofs of analytical results

Proof of Proposition 1. Suppose firm l has chosen $\overline{q}_{t,j,l}$ when the committee has selected strain j at time t. Using the inverse demand function in (7), we obtain the expected profit of firm l as follows:

$$\pi_{t,j,l}(\overline{q}_{t,j,l}) = E\left[\left\{\left(\frac{2c_2\overline{e}e_j(\hat{\theta})}{x\overline{e}+(1-x)e_j(\hat{\theta})} - c_3\right) - \frac{2c_2\overline{e}e_j(\hat{\theta})}{S\{x\overline{e}+(1-x)e_j(\hat{\theta})\}}\left(\sum_{\kappa=1}^{n}\widetilde{y}_{j,\kappa}\overline{q}_{t,j,\kappa}\right)\right\}\widetilde{y}_{j,l}\overline{q}_{t,j,l} - c_1\overline{q}_{t,j,l}\right]$$

$$= E\left[-\frac{2c_2\overline{e}e_j(\hat{\theta})}{S\{x\overline{e}+(1-x)e_j(\hat{\theta})\}}\widetilde{y}_{j,l}^2\overline{q}_{t,j,l}^2 + \left\{\left(\frac{2c_2\overline{e}e_j(\hat{\theta})}{x\overline{e}+(1-x)e_j(\hat{\theta})} - c_3\right)\widetilde{y}_{j,l} - \frac{2c_2\overline{e}e_j(\hat{\theta})}{S\{x\overline{e}+(1-x)e_j(\hat{\theta})\}}\left(\sum_{\kappa\neq l}\widetilde{y}_{j,\kappa}\overline{q}_{t,j,\kappa}\right)\widetilde{y}_{j,l} - c_1\right\}\overline{q}_{t,j,l}\right].$$

Let $\overline{q}_{t,j,-l}$ denote the number of embryonated eggs chosen by all firms other than firm l. Then the maximum expected profit of firm l is

$$\pi_{t,j,l}^*(\hat{\theta}) = \max_{0\le\overline{q}_{t,j,l}\le r(T-t)} \pi_{t,j,l}\left(\overline{q}_{t,j,l}, \overline{q}_{t,j,-l}\right)$$

$$= \max_{0\le\overline{q}_{t,j,l}\le r(T-t)} -\frac{2c_2\overline{e}e_j(\hat{\theta})}{S\{x\overline{e}+(1-x)e_j(\hat{\theta})\}}(\mu^2+\sigma_j^2)\overline{q}_{t,j,l}^2 + \left\{\left(\frac{2c_2\overline{e}e_j(\hat{\theta})}{x\overline{e}+(1-x)e_j(\hat{\theta})} - c_3\right)\mu - \frac{2c_2\overline{e}e_j(\hat{\theta})\mu^2(n-1)}{S\{x\overline{e}+(1-x)e_j(\hat{\theta})\}}\overline{q}_{t,j,-l} - c_1\right\}\overline{q}_{t,jl}.$$

Since $\pi_{t,j,l}$ is concave with respect to $\overline{q}_{t,j,l}$, we obtain a symmetric Nash equilibrium from the first order condition as follows:

$$\overline{q}_{t,j,l}(\hat{\theta}) = \min\left[\frac{S}{\mu(n+1+2\delta_j^2)}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2}\left(\frac{x}{e_j(\hat{\theta})}+\frac{1-x}{\overline{e}}\right)\right\},\ r(T-t)\right].$$

Then the expected total vaccine quantity is

$$E[\widetilde{q}_{t,j}(\hat{\theta})] = E\left[\sum_{l=1}^{n}\widetilde{y}_{j,l}\overline{q}_{t,j,l}(\hat{\theta})\right] = \min\left[\frac{nS}{n+1+2\delta_j^2}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2}\left(\frac{x}{e_j(\hat{\theta})}+\frac{1-x}{\overline{e}}\right)\right\},\ nr\mu(T-t)\right].$$

Equation (9) is an equivalent representation of the above equation where $t_j(\hat{\theta})$ solves $\frac{nS}{n+1+2\delta_j^2}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2}\left(\frac{x}{e_j(\hat{\theta})}+\frac{1-x}{\overline{e}}\right)\right\} = nr\mu(T-t_j(\hat{\theta}))$. The assumption that $q_j(\hat{\theta}) > 0$ for all j, x, and $\hat{\theta}$ is equivalent to

$$\bar{e} > \frac{c_3 + c_1\mu^{-1}}{2c_2} \text{ and } e_j(\hat{\theta}) > \frac{c_3 + c_1\mu^{-1}}{2c_2} \text{ for all } j \text{ and } \hat{\theta}. \tag{A1}$$

By observation of (9), $q_j(\hat{\theta})$ is increasing in $e_j(\hat{\theta})$ for $x > 0$ and in $\bar{e}$ for $x < 1$, but decreasing in δ_j. □

Proof of Proposition 2. **(a)** Substituting $x = 0$, $c^{(u)} = 2c_2\left(1 - \tilde{q}_{t,j}(\hat{\theta})/S\right)$, $E[\tilde{q}_{t,j}^2] = (E[\tilde{q}_{t,j}])^2 + Var[\tilde{q}_{t,j}] = (1+\delta_j^2 n^{-1})(E[\tilde{q}_{t,j}])^2$, and $\sum_{l=1}^{n} \bar{q}_{t,j,l} = \mu^{-1}E[\tilde{q}_{t,j}]$ into (8), we obtain after simplification

$$\begin{aligned} W_{t,j}(\hat{\theta}) &= -\frac{c_2 e_j(\hat{\theta})(1+\delta_j^2 n^{-1})}{S}(E[\tilde{q}_{t,j}])^2 + \left\{2c_2 e_j(\hat{\theta}) - (c_3 + c_1\mu^{-1})\right\}E[\tilde{q}_{t,j}] \\ &= c_2 E[\tilde{q}_{t,j}]\left(-\frac{1+\delta_j^2 n^{-1}}{S}E[\tilde{q}_{t,j}] + 2\right)e_j(\hat{\theta}) - (c_3 + c_1\mu^{-1})E[\tilde{q}_{t,j}]. \end{aligned} \tag{A2}$$

Substituting $x = 0$ into (9), $q_j = \frac{S}{1+n^{-1}+2\delta_j^2 n^{-1}}\left\{1 - \frac{c_3 + c_1\mu^{-1}}{2c_2\bar{e}}\right\}$. Because $\frac{c_3 + c_1\mu^{-1}}{2c_2\bar{e}} > 0$,

$$E[\tilde{q}_{t,j}] \le q_j < \frac{S}{1+n^{-1}+2\delta_j^2 n^{-1}} < \frac{S}{1+\delta_j^2 n^{-1}}. \tag{A3}$$

Therefore, the coefficient of $e_j(\hat{\theta})$ in (A2) is positive, so the result follows.

(b) Rearranging (A2) yields

$$W_{t,j}(\hat{\theta}) = -\frac{c_2 e_j(\hat{\theta})(1+\delta_j^2 n^{-1})}{S}\left[E[\tilde{q}_{t,j}] - \frac{S}{1+\delta_j^2 n^{-1}}\left\{1 - \frac{c_3 + c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}\right]^2 + \frac{Sc_2 e_j(\hat{\theta})}{1+\delta_j^2 n^{-1}}\left\{1 - \frac{c_3 + c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}^2.$$

From the assumption that vaccine supply in equilibrium is less than a socially-optimal quantity in expectation,

$$E[\tilde{q}_{t,j}] < \frac{S}{1+\delta_j^2 n^{-1}}\left\{1 - \frac{c_3 + c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}. \tag{A4}$$

Therefore, $W_{t,j}(\hat{\theta})$ is increasing in $E[\tilde{q}_{t,j}]$.

Using a chain rule, we obtain the first derivative of $W_{t,j}(\hat{\theta})$ with respect to δ_j^2 as follows:

$$\frac{\partial W_{t,j}(\hat{\theta})}{\partial \delta_j^2} = \left(\frac{\partial W_{t,j}(\hat{\theta})}{\partial \delta_j^2}\right)_{E[\tilde{q}_{t,j}]} + \frac{\partial E[\tilde{q}_{t,j}]}{\partial \delta_j^2}\frac{\partial W_{t,j}(\hat{\theta})}{\partial E[\tilde{q}_{t,j}]}$$

where $\left(\frac{\partial W_{t,j}(\hat{\theta})}{\partial \delta_j^2}\right)_{E[\tilde{q}_{t,j}]}$ represents the partial derivative of $W_{t,j}(\hat{\theta})$ with respect to δ_j^2 with $E[\tilde{q}_{t,j}]$ held constant. Since $\left(\frac{\partial W_{t,j}(\hat{\theta})}{\partial \delta_j^2}\right)_{E[\tilde{q}_{t,j}]} = -\frac{c_2 e_j(\hat{\theta}) n^{-1}(E[\tilde{q}_{t,j}])^2}{S} < 0$, $\frac{\partial E[\tilde{q}_{t,j}]}{\partial \delta_j^2} < 0$ by Proposition 1, and $\frac{\partial W_{t,j}(\hat{\theta})}{\partial E[\tilde{q}_{t,j}]} > 0$ as shown above, it follows that $\frac{\partial W_{t,j}(\hat{\theta})}{\partial \delta_j^2} < 0$. Therefore, $W_{t,j}(\hat{\theta})$ is decreasing in δ_j. □

Remark on the Proof of Proposition 2. The sufficient condition for the assumption in (A4) is:

$$e_j(\hat{\theta}) > \kappa_j \bar{e} \text{ where } \kappa_j \equiv \left\{ \frac{2c_2\bar{e}}{c_3 + c_1\mu^{-1}} \left(1 - \frac{1+\delta_j^2 n^{-1}}{1+n^{-1}+2\delta_j^2 n^{-1}} \right) + \frac{1+\delta_j^2 n^{-1}}{1+n^{-1}+2\delta_j^2 n^{-1}} \right\}^{-1} (< 1).$$

Define $\theta^{\max}$ and $\theta^{\min}$ such that each solves $e_2(\theta^{\max}) = \kappa_2 \bar{e}$ and $e_1(\theta^{\min}) = \kappa_1 \bar{e}$, respectively. Then the next Theorem 1 implies that this condition is equivalent to the condition that $\theta^b > \theta^{\min}$ and $\theta^a < \theta^{\max}$. It is also easy to see from (A2) that $W_{t,j}(\hat{\theta}) > 0$ for all t (< T), j, and $\hat{\theta}$.

Proof of Theorem 1. Define $W_j^U(\hat{\theta})$ and $W_{t,j}^C(\hat{\theta})$ such that $W_{t,j}(\hat{\theta})$ equals $W_j^U(\hat{\theta})$ when $t \le t_j(\hat{\theta})$ and $W_{t,j}^C(\hat{\theta})$ when $t > t_j(\hat{\theta})$. Substituting $x = 0$ and $E[\tilde{q}_{t,j}]$ in (9) into (A2), we get

$$W_j^U(\hat{\theta}) = \left\{ -\frac{c_2(1+\delta_j^2 n^{-1})}{S} q_j^2 + 2c_2 q_j \right\} e_j(\hat{\theta}) - (c_3 + c_1\mu^{-1}) q_j, \text{ and} \tag{A5}$$

$$W_{t,j}^C(\hat{\theta}) = \left[-\frac{c_2(1+\delta_j^2 n^{-1})}{S} \{nr\mu(T-t)\}^2 + 2c_2 nr\mu(T-t) \right] e_j(\hat{\theta}) - (c_3 + c_1\mu^{-1}) nr\mu(T-t). \tag{A6}$$

In addition to θ^a and θ^b defined in the main body, we define $t_3(\hat{\theta})$, $t_4(\hat{\theta})$, and θ^c as follows:

- $t_3(\hat{\theta})$ ($\in [t_1, t_2]$) solves $W_{t_3(\hat{\theta}),1}^C(\hat{\theta}) = W_2^U(\hat{\theta})$ for any $\hat{\theta}$.
- $t_4(\hat{\theta})$ ($\in [t_2, T]$) solves $W_{t_4(\hat{\theta}),1}^C(\hat{\theta}) = W_{t_4(\hat{\theta}),2}^C(\hat{\theta})$ for any $\hat{\theta}$.
- $\theta^c \equiv t_3^{-1}(t_2)$.

The proof proceeds in four steps by showing that: (a) θ^a exists in (0, 1) and θ^b exists with $\theta^b < \theta^a$; (b) $t_3(\hat{\theta})$ exists, and is increasing and concave in $\hat{\theta}$; (c) $t_4(\hat{\theta})$ exists, and is increasing and convex in $\hat{\theta}$; and (d) $t_3(\theta^b) = t_1$, $t_3(\theta^c) = t_4(\theta^c) = t_2$, and $t_4(\theta^a) = T$. Then the existence of θ^c in (θ^b, θ^a) immediately follows. The proof is completed by setting θ_t^S equal to θ^b for $t \in [0, t_1]$, $t_3^{-1}(t)$ for $t \in [t_1, t_2]$, and $t_4^{-1}(t)$ for $t \in [t_2, T]$.

(a) *The existence of θ^a in (0, 1)*: Since $e_1(\hat{\theta}) - e_2(\hat{\theta})$ is continuous and increasing in $\hat{\theta}$, $e_1(0) < e_2(0)$, and $e_1(1) > e_2(1)$, there exists a unique solution $\theta^a = (e_{22} - e_{12})/(e_{11} + e_{22} - e_{12} - e_{21})$ in (0, 1) that solves $e_1(\theta^a) = e_2(\theta^a)$.

The existence of θ^b less than θ^a: By Proposition 2, $W_1^U(\hat{\theta}) - W_2^U(\hat{\theta})$ is continuous and linearly increasing in $\hat{\theta}$. Since $e_1(\theta^a) = e_2(\theta^a)$ and $\delta_1 < \delta_2$, $W_1^U(\theta^a) > W_2^U(\theta^a)$ by Proposition 2. So, θ^b exists with $\theta^b < \theta^a$.

(b) By the same reasoning, $t_3(\hat{\theta})$ exists for any given $\hat{\theta}$. Rearranging '$W_{t_3(\hat{\theta}),1}^C(\hat{\theta}) = W_2^U(\hat{\theta})$' yields

$$\left[nr\mu(T-t_3(\hat{\theta}))\right]^2 - \frac{2S}{1+\delta_1^2\, n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2e_1(\hat{\theta})}\right\}\left[nr\mu(T-t_3(\hat{\theta}))\right] + \frac{S}{c_2e_1(\hat{\theta})(1+\delta_1^2\, n^{-1})}W_2^U(\hat{\theta}) = 0\,.$$

Since $n\,r\mu(T-t_3(\hat{\theta})) < \frac{S}{1+\delta_1^2\, n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2e_1(\hat{\theta})}\right\}$ by (A4), $nr\mu(T-t_3(\hat{\theta}))$ is a smaller root of the above quadratic equation.

Differentiating both sides of '$W^C_{t_3(\hat{\theta}),1}(\hat{\theta}) = W_2^U(\hat{\theta})$' with respect to $\hat{\theta}$ by using a chain rule, we obtain

$$\left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}}\right|_{t=t_3(\hat{\theta})} + \frac{dt_3(\hat{\theta})}{d\hat{\theta}}\left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t}\right|_{t=t_3(\hat{\theta})} = \frac{dW_2^U(\hat{\theta})}{d\hat{\theta}}. \tag{A7}$$

Rearranging (A7) gives

$$\frac{dt_3(\hat{\theta})}{d\hat{\theta}} = \left(\frac{dW_2^U(\hat{\theta})}{d\hat{\theta}} - \left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}}\right|_{t=t_3(\hat{\theta})}\right) \Big/ \left(\left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t}\right|_{t=t_3(\hat{\theta})}\right). \tag{A8}$$

Since $\frac{dW_2^U(\hat{\theta})}{d\hat{\theta}} < 0$, $\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}} > 0$, and $\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t} < 0$ for all t and $\hat{\theta}$ by Proposition 2, it follows from (A8) that $\frac{dt_3(\hat{\theta})}{d\hat{\theta}} > 0$ for all $\hat{\theta}$.

Differentiating both sides of (A7) with respect to $\hat{\theta}$ by using a chain rule, we obtain

$$\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}^2}\right|_{t=t_3(\hat{\theta})} + \frac{dt_3(\hat{\theta})}{d\hat{\theta}}\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t\partial\hat{\theta}}\right|_{t=t_3(\hat{\theta})} + \frac{d^2t_3(\hat{\theta})}{d\hat{\theta}^2}\left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t}\right|_{t=t_3(\hat{\theta})} + \frac{dt_3(\hat{\theta})}{d\hat{\theta}}\left(\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t\partial\hat{\theta}}\right|_{t=t_3(\hat{\theta})} + \frac{dt_3(\hat{\theta})}{d\hat{\theta}}\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t^2}\right|_{t=t_3(\hat{\theta})}\right) = \frac{d^2W_2^U(\hat{\theta})}{d\hat{\theta}^2}. \tag{A9}$$

Rearranging (A9) gives

$$\frac{d^2t_3(\hat{\theta})}{d\hat{\theta}^2} = \left\{\frac{d^2W_2^U(\hat{\theta})}{d\hat{\theta}^2} - \left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}^2}\right|_{t=t_3(\hat{\theta})} - 2\frac{dt_3(\hat{\theta})}{d\hat{\theta}}\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t\partial\hat{\theta}}\right|_{t=t_3(\hat{\theta})} - \left(\frac{dt_3(\hat{\theta})}{d\hat{\theta}}\right)^2\left.\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t^2}\right|_{t=t_3(\hat{\theta})}\right\} \Big/ \left.\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t}\right|_{t=t_3(\hat{\theta})}. \tag{A10}$$

By Proposition 2, $\frac{d^2W_2^U(\hat{\theta})}{d\hat{\theta}^2} = \frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial\hat{\theta}^2} = 0$ for all t and $\hat{\theta}$. From (A6), we get

$$\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t^2} = -\frac{2c_2nr^2\mu^2(n+\delta_1^2)e_1(\hat{\theta})}{S} < 0. \tag{A11}$$

Since $e_{11} - e_{12} > 0$ and $n\,r\mu(T-t) < \frac{S}{1+\delta_1^2 n^{-1}}$ by (A3),

$$\frac{\partial^2 W^C_{t,1}(\hat{\theta})}{\partial t\partial\hat{\theta}} = 2c_2nr\mu(e_{11}-e_{12})\left\{-1+\frac{r\mu(n+\delta_1^2)(T-t)}{S}\right\} < 0. \tag{A12}$$

Combining $\frac{dt_3(\hat{\theta})}{d\hat{\theta}} > 0$ and $\frac{\partial W^C_{t,1}(\hat{\theta})}{\partial t} < 0$ with these results, it follows that $\frac{d^2t_3(\hat{\theta})}{d\hat{\theta}^2} < 0$ for all $\hat{\theta}$.

(c) Solving $W^{C}_{t_4(\hat{\theta}),1}(\hat{\theta}) = W^{C}_{t_4(\hat{\theta}),2}(\hat{\theta})$, we get

$$t_4(\hat{\theta}) = T - \frac{2S\{e_1(\hat{\theta}) - e_2(\hat{\theta})\}}{r\mu\{(n+\delta_1^2)e_1(\hat{\theta}) - (n+\delta_2^2)e_2(\hat{\theta})\}}. \tag{A13}$$

Define $e_{1/2}(\hat{\theta}) \equiv e_1(\hat{\theta}) / e_2(\hat{\theta})$. Since $e_1(\hat{\theta}) < e_2(\hat{\theta})$ for $\hat{\theta} \in [\theta^b, \theta^a]$, $0 < e_{1/2}(\hat{\theta}) < 1$. By simple algebra, we get $e^{'}_{1/2}(\hat{\theta}) > 0$ and $e^{''}_{1/2}(\hat{\theta}) > 0$ for all $\hat{\theta}$. Using $\delta_1 < \delta_2$, we then obtain

$$\frac{dt_4(\hat{\theta})}{d\hat{\theta}} = \frac{2S}{r\mu} \cdot \frac{(\delta_2^2 - \delta_1^2)e^{'}_{1/2}(\hat{\theta})}{\{(n+\delta_1^2)e_{1/2}(\hat{\theta}) - (n+\delta_2^2)\}^2} > 0, \text{ and}$$

$$\frac{d^2t_4(\hat{\theta})}{d\hat{\theta}^2} = \frac{2S}{r\mu} \cdot \frac{(\delta_2^2 - \delta_1^2)\,[e^{''}_{1/2}(\hat{\theta})\{(n+\delta_1^2)e_{1/2}(\hat{\theta}) - (n+\delta_2^2)\} - 2(n+\delta_1^2)\{e^{'}_{1/2}(\hat{\theta})\}^2]}{\{(n+\delta_1^2)e_{1/2}(\hat{\theta}) - (n+\delta_2^2)\}^3} > 0 \text{ for all } \hat{\theta}.$$

(d) By the definition of θ^b, $W_1^U(\theta^b) = W_2^U(\theta^b)$. By the definition of t_1, $W_1^U(\hat{\theta}) = W^{C}_{t_1,1}(\hat{\theta})$ for all $\hat{\theta}$. Thus, $W_1^U(\theta^b) = W^{C}_{t_1,1}(\theta^b) = W_2^U(\theta^b)$, hence $t_3(\theta^b) = t_1$ by the definition of $t_3(\hat{\theta})$.

By the definition of t_2, $W_2^U(\hat{\theta}) = W^{C}_{t_2,2}(\hat{\theta})$ for all $\hat{\theta}$. By the definition of $t_3(\hat{\theta})$, $W^{C}_{t_3(\hat{\theta}),1}(\hat{\theta}) = W_2^U(\hat{\theta})$ for all $\hat{\theta}$. Since $t_2 = t_3(\theta^c)$ by the definition of θ^c, $W^{C}_{t_3(\theta^c),1}(\theta^c) = W^{C}_{t_3(\theta^c),2}(\theta^c)$, hence $t_3(\theta^c) = t_4(\theta^c)$ by the definition of $t_4(\hat{\theta})$.

Substituting $\hat{\theta} = \theta^a$ into (A13) and using $e_1(\theta^a) = e_2(\theta^a)$, we obtain $t_4(\theta^a) = T$. □

***Corollary* A1.**

(a) For all $\hat{\theta}$ and $t > t_1$, $W_1^U(\hat{\theta}) > W^{C}_{t,1}(\hat{\theta}) > W^{C}_{t+1,1}(\hat{\theta})$ and $W_1^{U'}(\hat{\theta}) > W^{C'}_{t,1}(\hat{\theta}) > W^{C'}_{t+1,1}(\hat{\theta}) > 0$.

(b) For all $\hat{\theta}$ and $t > t_2$, $W_2^U(\hat{\theta}) > W^{C}_{t,2}(\hat{\theta}) > W^{C}_{t+1,2}(\hat{\theta})$ and $W_2^{U'}(\hat{\theta}) < W^{C'}_{t,2}(\hat{\theta}) < W^{C'}_{t+1,2}(\hat{\theta}) < 0$.

(Note: we use the notation of '$W^{'}_{t}(\hat{\theta}) \equiv \partial W_t(\hat{\theta}) / \partial\hat{\theta}$' in the remainder of Appendix.)

Proof of Corollary A1. **(a)** Since $W_{t,j}(\hat{\theta})$ is increasing in $E[\tilde{q}_{t,j}]$ for all $\hat{\theta}$ and $q_1 > nr\mu(T-t) > nr\mu(T-t-1)$ for $t > t_1$, it follows that $W_1^U(\hat{\theta}) > W^{C}_{t,1}(\hat{\theta}) > W^{C}_{t+1,1}(\hat{\theta})$. Since $W_1^{U'}(\hat{\theta}) = W^{C'}_{t_1,1}(\hat{\theta})$, $\frac{\partial W^{C}_{t,1}(\hat{\theta})}{\partial\hat{\theta}} > 0$, and $\frac{\partial^2 W^{C}_{t,1}(\hat{\theta})}{\partial t\partial\hat{\theta}} < 0$ by (A12) for all $\hat{\theta}$ and any $t > t_1$, $W_1^{U'}(\hat{\theta}) > W^{C'}_{t,1}(\hat{\theta}) > W^{C'}_{t+1,1}(\hat{\theta}) > 0$.

(b) A similar argument holds for *s*2. □

Proof of Theorem 2. For any *t*, both $W_t(\hat{\theta})$ and $EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$ are continuous in $\hat{\theta}$. Since $\Theta_{t+1}(\hat{\theta}) = \hat{\theta}$ with probability 1 when $\hat{\theta}$ approaches $\omega = 0$ or 1,

$$W_t(\omega) \geq W_{t+1}(\omega) = \lim_{\hat{\theta}\to\omega} EW_{t+1}(\Theta_{t+1}(\hat{\theta})) = \lim_{\hat{\theta}\to\omega} EV_{t+1}(\Theta_{t+1}(\hat{\theta})). \tag{A14}$$

Since the expectation or maximum of convex functions is also convex, by a simple induction argument, $V_t(\hat{\theta})$ is convex in $\hat{\theta}$ for all *t*. Combining the convexity of $EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$ with Corollary A1, we get

$$W'_{t,2}(\hat{\theta}) \le \frac{\partial EV_{t+1}(\Theta_{t+1}(\hat{\theta}))}{\partial \hat{\theta}} \le W'_{t,1}(\hat{\theta}) \text{ for any } \hat{\theta} \in (0, 1) \text{ at which } \frac{\partial EV_{t+1}(\Theta_{t+1}(\hat{\theta}))}{\partial \hat{\theta}} \text{ exists.} \quad \text{(A15)}$$

Now, if $W_t(\theta_t^S) \ge EV_{t+1}(\Theta_{t+1}(\theta_t^S))$, then $W_t(\hat{\theta}) \ge EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$ for all $\hat{\theta}$ by (A14) and (A15); therefore, $\theta_t^* = \theta_t^{**} = \theta_t^S$ such that $V_t(\hat{\theta}) = W_{t,1}(\hat{\theta})$ if $\hat{\theta} \ge \theta_t^*$, and $V_t(\hat{\theta}) = W_{t,2}(\hat{\theta})$ if $\hat{\theta} \le \theta_t^{**}$. On the other hand, if $W_t(\theta_t^S) < EV_{t+1}(\Theta_{t+1}(\theta_t^S))$, then there exists a pair of numbers θ_t^* and θ_t^{**} with $\theta_t^* < \theta_t^S < \theta_t^{**}$ by (A14) and (A15) such that $V_t(\hat{\theta}) = W_{t,1}(\hat{\theta})$ if $\theta \ge \theta_t^*$, $V_t(\hat{\theta}) = EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$ if $\theta_t^* < \hat{\theta} < \theta_t^{**}$, and $V_t(\hat{\theta}) = W_{t,2}(\hat{\theta})$ if $\hat{\theta} \le \theta_t^{**}$. Therefore, there exists a pair of threshold numbers θ_t^* and θ_t^{**} with $0 \le \theta_t^* \le \theta_t^S \le \theta_t^{**} \le 1$.

In order to investigate how θ_t^* and θ_t^{**} change over *t*, we divide the time horizon into three intervals: $I_1 \equiv [0, t_1 - 1]$, $I_2 \equiv (t_1 - 1, t_2 - 1]$, and $I_3 \equiv (t_2 - 1, T - 1]$. Note that the non-monotonicity of θ_t^{**} is established by one counter-example where θ_t^{**} is not monotone in *t*. The computational examples shown in Figures 4-6 serve this purpose.

For any $t \in I_1$, θ_t^S does not change in *t* and $W_{t+1}(\hat{\theta}) = W_t(\hat{\theta})$ for all $\hat{\theta}$. Therefore, for all $\hat{\theta}$,

$$EV_{t+1}(\Theta_{t+1}(\hat{\theta})) \ge EW_{t+1}(\Theta_{t+1}(\hat{\theta})) \ge W_{t+1}(\hat{\theta}) = W_t(\hat{\theta}) \quad \text{(A16)}$$

where the first inequality follows from the definition of V_{t+1} and the second inequality follows from Jensen's inequality. Therefore, $\theta_t^* = 0$ and $\theta_t^{**} = 1$ for all $t \in I_1$.

For any $t \in I_2$, as θ_t^S increases in *t* by Theorem 1, (A16) continues to hold for any $\hat{\theta} \le \theta_t^S$, hence $\theta_t^* = 0$ for all $t \in I_2$. Thus, $W_t(\hat{\theta})$ crosses $EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$ at one point $\hat{\theta} = \theta_t^{**}$ greater than θ_t^S.

For any $t \in I_3$, by Corollary A1, $W_t(\hat{\theta}) > W_{t+1}(\hat{\theta})$ for all $\hat{\theta}$, so $W_t(\hat{\theta})$ may or may not cross $EV_{t+1}(\Theta_{t+1}(\hat{\theta}))$. At $T-1$, $V_{T-1}(\hat{\theta}) = \max\{W_{T-1}(\hat{\theta}), EV_T(\Theta_T(\hat{\theta}))\} = W_{T-1}(\hat{\theta})$ because $W_{T-1}(\hat{\theta}) > 0$ and $V_T(\hat{\theta}) = 0$ for all $\hat{\theta}$; hence, $\theta_{T-1}^* = \theta_{T-1}^{**} = \theta_{T-1}^S$. Thus, both θ_t^* and θ_t^{**} eventually converge to θ_t^S.

In the rest of the proof, we show that θ_t^* is increasing in $t \in I_3$. Let $\bar{t}$ $(\in I_3)$ denote the largest *t* such that $\theta_t^* < \theta_t^S$. Then, $\theta_t^* = \theta_t^{**} = \theta_t^S$ for all $t > \bar{t}$, hence θ_t^* is increasing in $t > \bar{t}$. In order to show that $\theta_t^* < \theta_{t+1}^*$ for $t \le \bar{t}$, it suffices to show that if it is optimal to select *s*2 with the estimate $\hat{\theta}$ at *t*, then it is strictly optimal to select *s*2 with the same estimate $\hat{\theta}$ at *t*+1, *i.e.*

$$\{W_{t,2}(\hat{\theta}) \geq W_{t,1}(\hat{\theta}) \text{ and } W_{t,2}(\hat{\theta}) \geq EV_{t+1}(\Theta_{t+1}(\hat{\theta}))\} \Rightarrow \{W_{t+1,2}(\hat{\theta}) > W_{t+1,1}(\hat{\theta}) \text{ and } W_{t+1,2}(\hat{\theta}) > EV_{t+2}(\Theta_{t+2}(\hat{\theta}))\}. \quad \text{(A17)}$$

Since θ_t^S is increasing in t, $W_{t,2}(\hat{\theta}) \geq W_{t,1}(\hat{\theta}) \Rightarrow W_{t+1,2}(\hat{\theta}) > W_{t+1,1}(\hat{\theta})$. Thus, (A17) holds if the following condition is satisfied for all $\hat{\theta}$ and t in $(t_2 - 1, \bar{t}\,]$:

$$EV_{t+1}(\Theta_{t+1}(\hat{\theta})) - EV_{t+2}(\Theta_{t+2}(\hat{\theta})) - W_{t,2}(\hat{\theta}) + W_{t+1,2}(\hat{\theta}) > 0. \quad \text{(A18)}$$

Because $\Theta_{t+1}(\hat{\theta})$ is riskier than $\Theta_{t+2}(\hat{\theta})$ in the sense of second-order stochastic dominance and V_{t+1} is convex, $EV_{t+1}(\Theta_{t+1}(\hat{\theta})) \geq EV_{t+1}(\Theta_{t+2}(\hat{\theta}))$. Thus, (A18) is implied by

$$EV_{t+1}(\Theta_{t+2}(\hat{\theta})) - EV_{t+2}(\Theta_{t+2}(\hat{\theta})) - W_{t,2}(\hat{\theta}) + W_{t+1,2}(\hat{\theta}) > 0. \quad \text{(A19)}$$

Since $\partial^2 W_{t,2}^C(\hat{\theta})/\partial t^2 < 0$ for any fixed $\hat{\theta}$ from (A11), $W_{t+1,2}(\hat{\theta}) - W_{t+2,2}(\hat{\theta}) > W_{t,2}(\hat{\theta}) - W_{t+1,2}(\hat{\theta})$, so (A19) is implied by

$$EV_{t+1}(\Theta_{t+2}(\hat{\theta})) - EV_{t+2}(\Theta_{t+2}(\hat{\theta})) - W_{t+1,2}(\hat{\theta}) + W_{t+2,2}(\hat{\theta}) \geq 0. \quad \text{(A20)}$$

Because $W_{t,2}(\hat{\theta})$ is a linear function of $\hat{\theta}$ for all t from Proposition 2, $W_{t+1,2}(\hat{\theta}) = EW_{t+1,2}(\Theta_{t+2}(\hat{\theta}))$ and $W_{t+2,2}(\hat{\theta}) = EW_{t+2,2}(\Theta_{t+2}(\hat{\theta}))$. Thus, (A20) can be rewritten as

$$E[\, V_{t+1}(\Theta_{t+2}(\hat{\theta})) - V_{t+2}(\Theta_{t+2}(\hat{\theta})) - W_{t+1,2}(\Theta_{t+2}(\hat{\theta})) + W_{t+2,2}(\Theta_{t+2}(\hat{\theta}))\,] \geq 0.$$

This is true if for any x and $t \in I_3$,

$$\{V_{t+1}(x) - W_{t+1,2}(x)\} - \{V_{t+2}(x) - W_{t+2,2}(x)\} \geq 0. \quad \text{(A21)}$$

We prove (A21) using an induction argument on t. For any $t \geq \bar{t}$, $V_{t+1}(x) = W_{t+1}(x)$ and $V_{t+2}(x) = W_{t+2}(x)$ for all x, so (A21) becomes

$$\{W_{t+1}(x) - W_{t+1,2}(x)\} - \{W_{t+2}(x) - W_{t+2,2}(x)\} \geq 0. \quad \text{(A22)}$$

For $x \leq \theta_{t+1}^S$, $W_{t+1}(x) = W_{t+1,2}(x)$ and $W_{t+2}(x) = W_{t+2,2}(x)$, so (A22) holds with equality. For $\theta_{t+1}^S \leq x \leq \theta_{t+2}^S$, $W_{t+1}(x) = W_{t+1,1}(x) \geq W_{t+1,2}(x)$ and $W_{t+2}(x) = W_{t+2,2}(x)$, so (A22) holds with inequality. For $x \geq \theta_{t+2}^S$, $W_{t+1}(x) = W_{t+1,1}(x)$ and $W_{t+2}(x) = W_{t+2,1}(x)$. If we define $f_{t+1}(x) \equiv \{W_{t+1,1}(x) - W_{t+1,2}(x)\} - \{W_{t+2,1}(x) - W_{t+2,2}(x)\}$, $f_{t+1}(\theta_{t+2}^S) > 0$ and $\partial f_{t+1}(x)/\partial x > 0$ for all x given $t \in I_3$ by Corollary A1. Therefore, (A22) holds for any x and $t \in I_3$.

Now, suppose (A21) holds for any fixed $t+1 \in (t_2, \bar{t}\,]$. This implies that (A18) holds at $t+1$, hence $\theta_{t+1}^* < \theta_{t+2}^*$. We show that (A21) also holds at t in each of the following cases: (Case 1) $\theta_{t+1}^{**} \geq \theta_{t+2}^{**}$ and (Case 2) $\theta_{t+1}^{**} < \theta_{t+2}^{**}$. (Recall that the upper threshold θ_t^{**} is not necessarily monotone in t.)

(Case 1) $\theta^{**}_{t+1} \geq \theta^{**}_{t+2}$:

(i) For $x \leq \theta^{*}_{t+1}$, $V_{t+1}(x) = W_{t+1,2}(x)$ and $V_{t+2}(x) = W_{t+2,2}(x)$. So, (A21) holds with equality.

(ii) For $x \in [\theta^{*}_{t+1}, \theta^{*}_{t+2}]$, $V_{t+1}(x) = EV_{t+2}(\Theta_{t+2}(x)) \geq W_{t+1,2}(x)$, and $V_{t+2}(x) = W_{t+2,2}(x)$. So, (A21) holds with inequality.

(iii) For $x \in [\theta^{*}_{t+2}, \theta^{**}_{t+2}]$, $V_{t+1}(x) = EV_{t+2}(\Theta_{t+2}(x))$, and $V_{t+2}(x) = EV_{t+3}(\Theta_{t+3}(x))$. By the induction hypothesis of (A18), (A21) holds with strict inequality.

(iv) For $x \in [\theta^{**}_{t+2}, \theta^{**}_{t+1}]$, $V_{t+1}(x) = EV_{t+2}(\Theta_{t+2}(x)) \geq W_{t+1,1}(x)$, and $V_{t+2}(x) = W_{t+2,1}(x)$. So, (A21) follows from (A22).

(v) For $x \geq \theta^{**}_{t+1}$, $V_{t+1}(x) = W_{t+1,1}(x)$, and $V_{t+2}(x) = W_{t+2,1}(x)$. The result follows from (A22).

(Case 2) $\theta^{**}_{t+1} < \theta^{**}_{t+2}$:

The proof of only (iv) differs from (Case 1) as follows.

(iv') For $x \in [\theta^{**}_{t+1}, \theta^{**}_{t+2}]$, $V_{t+1}(x) = W_{t+1,1}(x)$, and $V_{t+2}(x) = EV_{t+3}(\Theta_{t+3}(x))$. If we define $g_{t+1}(x) \equiv \{W_{t+1,1}(x) - EV_{t+3,2}(\Theta_{t+3}(x))\} - \{W_{t+1,2}(x) - W_{t+2,2}(x)\}$, it follows from the induction hypothesis of (A18) that $g_{t+1}((\theta^{**}_{t+1}) > 0$ because $W_{t+1,1}(\theta^{**}_{t+1}) = EV_{t+2}(\Theta_{t+2}(\theta^{**}_{t+1}))$. By Corollary A1 and (A15), $W'_{t+1,1}(x) > W'_{t+2,1}(x) \geq \partial EV_{t+3}(\Theta_{t+3}(x))/\partial x$ and $W'_{t+1,2}(x) < W'_{t+2,2}(x)$, hence $\partial g_{t+1}(x)/\partial x > 0$ for all x given $t \in I_3$. Therefore, $g_{t+1}(x) > 0$ for any $x \geq \theta^{**}_{t+1}$. □

Proof of Proposition 3. **(a)** Similar to (A2), we obtain from (8):

$$W_{t,j}(\hat{\theta}) = -\frac{c_2 e_j(\hat{\theta})(1+\delta_j^2 n^{-1})}{S}\{E[\tilde{q}_{t,j}(\hat{\theta})]\}^2 + \left\{2c_2 e_j(\hat{\theta}) - (c_3 + c_1\mu^{-1})\right\}E[\tilde{q}_{t,j}(\hat{\theta})]. \tag{A23}$$

Define $W^U_j(\hat{\theta})$ and $W^C_{t,j}(\hat{\theta})$ as in the case where $x = 0$. Substituting $x = 1$ and $E[\tilde{q}_{t,j}(\hat{\theta})]$ in (9) into (A23), we obtain after simplification

$$W^U_j(\hat{\theta}) = \frac{Sc_2 n(n+2+3\delta_j^2)}{(n+1+2\delta_j^2)^2}\left\{e_j(\hat{\theta}) + \left(\frac{c_3 + c_1\mu^{-1}}{2c_2}\right)^2 \frac{1}{e_j(\hat{\theta})} - \frac{c_3 + c_1\mu^{-1}}{c_2}\right\}, \text{ and} \tag{A24}$$

$$W^C_{t,j}(\hat{\theta}) = c_2 nr\mu(T-t)\left\{2 - \frac{nr\mu(T-t)(1+\delta_j^2 n^{-1})}{S}\right\}e_j(\hat{\theta}) - nr\mu(T-t)(c_3 + c_1\mu^{-1}). \tag{A25}$$

It is easy to see that W^U_j is increasing *convexly* in $e_j(\hat{\theta})$ and that $W^C_{t,j}$ is increasing *linearly* in $e_j(\hat{\theta})$ because $2 - \frac{nr\mu(T-t)(1+\delta_j^2 n^{-1})}{S} > 0$ by (A3).

(b) Rearranging (A23) yields

$$W_{t,j}(\hat{\theta}) = -\frac{c_2 e_j(\hat{\theta})(1+\delta_j^2 n^{-1})}{S}\left[E[\tilde{q}_{t,j}(\hat{\theta})] - \frac{S}{1+\delta_j^2 n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}\right]^2 + \frac{Sc_2 e_j(\hat{\theta})}{1+\delta_j^2 n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}^2.$$

Since $E[\tilde{q}_{t,j}(\hat{\theta})] \le q_j(\hat{\theta}) = \frac{S}{1+n^{-1}+2\delta_j^2 n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\} < \frac{S}{1+\delta_j^2 n^{-1}}\left\{1-\frac{c_3+c_1\mu^{-1}}{2c_2 e_j(\hat{\theta})}\right\}$, $W_{t,j}(\hat{\theta})$ is increasing in $E[\tilde{q}_{t,j}(\hat{\theta})]$ when others are held constant.

In order to show that W_j^U is decreasing in δ_j, it suffices to show that $\frac{n(n+2+3\delta_j^2)}{(n+1+2\delta_j^2)^2}$ is decreasing in δ_j because $e_j(\hat{\theta}) + \left(\frac{c_3+c_1\mu^{-1}}{2c_2}\right)^2 \frac{1}{e_j(\hat{\theta})} - \frac{c_3+c_1\mu^{-1}}{c_2} > 0$ by (A1). Since the first derivative of $\frac{n(n+2+3\delta_j^2)}{(n+1+2\delta_j^2)^2}$ with respect to δ_j^2 is negative for $\delta_j^2 > 0$, the result follows. By observation of (A25), $W_{t,j}^C$ is decreasing in δ_j.

The proof of Theorem 1 proceeds similarly to the case where $x = 0$, hence is omitted. □

B. Model calibration for vaccine demand

We use the expected costs of infection in Table A1 for different age and health-risk groups estimated by Weycker et al. (2005). Using the different cost components of Weycker et al. (2005), we compute in Table A1 vaccine coverage rates when vaccine price p = \$9. We compute the total vaccine coverage, assuming the expected cost of infection is uniformly distributed within the entire population. Similarly, we compute the vaccine coverage of each subgroup, assuming the expected cost of infection is uniformly distributed within the group. Comparing the estimated coverage with the actual coverage shown in Table A2, we find that the vaccine coverage estimated using 'c_2 = **a** + **b** + **c**' is much higher than the actual coverage – in particular, the coverage of 0~64 years is overestimated significantly while the coverage of 65~ years is underestimated. The underestimation of the coverage of 65~ years is due to our assumption of the same vaccine price for all groups although the price for this group is actually lower due to Medicare reimbursement. The overestimation of the coverage of 0~64 years may be because individuals do not internalize their indirect costs of infection due to work loss (**b**) and death (**c**). To reduce this discrepancy, we set **c** = 0 and proportionally reduce **b** until the total demand approaches 103 million doses, the amount the industry actually produced before the British regulatory authorities prohibited the sale of Chiron's vaccine. We set the expected cost of infection among the population (c_2) to \$18, which is the weighted average of '**a** + 0.5**b**' among the entire population. This results in q_2 = 101 million doses and p = \$9 in

equilibrium when $x = 0$. Note that the actual coverage during the 2004-2005 season turned out to be much lower due to the Chiron's production failure. The mean charge for vaccination was $9.64 for the 1998-1999 season, which was higher than the Medicare reimbursement of $8.50 (Nichol 2001). The Medicare reimbursement for the 2002-2003 season was approximately $7.61 (source: http://www.acponline.org/hpp/reimbursement.htm).

Table A1: Vaccine coverage rates estimated by the model for different values of c_2.

	0~4 years	**5~18 years**		**19~64 years**		**65~ years**	**Total**
		Healthy	Risky	Healthy	Risky		
Cost data from Weycker et al. (2005) ($)							
Exp. healthcare cost (**a**)	36	7	19	4	14	19	41
Exp. work loss cost (**b**)	19	33	33	19	19	4	28
Exp. future lost earnings (**c**)	4	2	38	9	69	0	18
Population (%)	7%	16%	2%	52%	11%	12%	100%
Vaccine coverage estimated by the model							
$c_2 = \mathbf{a} + \mathbf{b} + \mathbf{c}$	67%	72%	87%	61%	88%	48%	71%
$c_2 = \mathbf{a} + \mathbf{b}$	63%	70%	77%	47%	64%	48%	57%
$c_2 = \mathbf{a} + 0.5\mathbf{b}$	47%	50%	67%	20%	49%	44%	34%

Note: *Population (%)* is based on the U.S. population estimates in 2005 (U.S. Census Bureau 2006) and the proportion of healthy (risky) population of 90% (10%) (Weycker et al. 2005).

Table A2: Vaccine coverage rates from surveys.

	6~23 months	**2~17 years**		**18~64 years**		**65~ years**	**Total**
		Healthy	Risky	Healthy	Risky		
2003-2004				21%	31%	65%	28%
2004-2005	48%	12%	35%	11%	22%	60%	19%

Note 1: 6 months ~ 17 years (CDC 2005), 18 ~ years (National Health Interview Survey 2005), and total (Strikas 2005).

Note 2: Surveys for pediatric groups began from the 2004-2005 season.

Note 3: Vaccine coverage in the 2004-2005 season is lower than that in a typical year because of a significant vaccine shortage.

References

Anderson, R. M., May, R. M. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, New York.

Andreasen, V., J. Lin, S. A. Levin. 1997. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology* **35** 825-842.

Bailey, N. T. J. 1975. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, London, UK.

Bayus, B. L. 1997. Speed-to-market and product performance trade-offs. *Journal of Product Innovation Management* **14** 485-497.

Bearden, J. N., A. Rapoport, R. O. Murphy. 2006. Sequential observation and selection with rank-dependent payoffs: an experimental study. *Management Science* **52** (9) 1437-1449.

Brito, D. L., E. Sheshinski, M. D. Intriligator. 1991. Externalities and compulsory vaccinations. *Journal of Public Economics* **45** 69-90.

Castillo-Chavez, C., H. W. Hethcote, V. Andreasen, S. A. Levin, W. M. Liu. 1989. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *Journal of Mathematical Biology* **27** 233-258.

Centers for Disease Control and Prevention. 2005. Estimated influenza vaccination coverage among adults and children – United States, September 1, 2004 – January 31, 2005. *MMWR* April 1; **54** (12) 304-307.

Centers for Disease Control and Prevention. 2007. Influenza (Flu). Retrieved October 20, 2007, http://www.cdc.gov/ flu.

Chick, S. E., H. Mamani, D. Simchi-Levi. 2006. Supply chain coordination and influenza vaccination. Working paper, INSEAD.

Cohen, M. A., J. Eliashberg, T. Ho. 1996. New product development: The performance and time-to-market tradeoff. *Management Science* **42**(2) 173-186.

Corbett, C. J., S. Deo. 2006. Cournot competition under yield uncertainty: the case of the influenza vaccine market. Working paper, UCLA Anderson School of Management.

Earn, D., J. Dushoff, S. A. Levin. 2002. Ecology and evolution of the flu. *TRENDS in Ecology & Evolution* **17**(7) 334-340.

Federgruen, A., N. Yang. 2007. Selecting a portfolio of suppliers under demand and supply risks. *Operations Research* (forthcoming).

GAO. 2001. Flu vaccine: Supply problems heighten need to ensure access for high-risk peopl*e*. *Report To Congressional Requesters*.

Gerdil, C. 2003. The annual production cycle for influenza vaccine. *Vaccine* **21** 1776-1779.

Hill, A. N., I. M. Longini. 2003. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences* **181** 85-106.

Jefferson, T., D. Rivetti, A. Rivetti, M. Rudin, C. Di Pietrantonj, V. Demicheli. 2005. Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review. *Lancet* **366** 1165-1174.

Kornish, L. J., R. L. Keeney. 2006. Repeated commit-or-defer decisions with a deadline: the influenza vaccine composition. *Operations Research* (forthcoming).

Levine, A. J. 1992. *Viruses*. W.H. Freeman & Co. New York.

Matthews, J. T. 2006. Egg-based production of influenza vaccine: 30 years of commercial experience. *The Bridge* **36**(3) 17-24.

McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management Science* **31** 1372-1389.

Mizuta, K., T. Itagaki, C. Abiko, T. Murata, T. Takahashi, S. Murayama. 2004. Epidemics of two Victoria and Yamagata influenza B lineages in Yamagata, Japan. *Epidemiol*. *Infect*. **132** 721-726.

Morgan, L. O., R. M. Morgan, W. L. Moore. 2001. Quality and time-to-market trade-offs when there are multiple product generations. *Manufacturing & Service Operations Management* **3**(2) 89-104.

Morris, W. T. 1968. *Management Science: A Bayesian Introduction*. Prentice-Hall, New Jersey.

Nochol, K. L. 2001. Cost-benefit analysis of a strategy to vaccinate healthy working adults against influenza. *Arch Intern Med* **161** 749-759.

Nichol, K. L. 2003. The efficacy, effectiveness, and cost-effectiveness of inactivated influenza virus vaccines. *Vaccine* **21** 1769-1775.

Offit, P. A. 2005. Why are pharmaceutical companies gradually abandoning vaccines? *Health Affairs* **24** (3) 622-630.

O'Mara, D., K. Fukuda, J. A. Singleton. 2003. Influenza vaccine: ensuring timely and adequate supply. *Infect. Med.* **20**(11) 548-554.

Özer, Ö., O. Uncu. 2006. Optimizing dynamic time-to-market and production decisions for a component supplier. Working paper, Stanford University.

Philipson, T. 2003. Economic epidemiology and infectious diseases. *Handbook of Health Economics*. Vol. 1B 1761-1799 (eds. Culyer A.J., J.P. Newhouse).

Souza, G. C., B. L. Bayus, H. M. Wagner. 2004. New-product strategy and industry clockspeed. *Management Science* **50**(4) 537-549.

Strikas, R. A. 2005. Driving increased influenza vaccine uptake (extending the vaccination season). *The 2005 National Influenza Summit*, Chicago, IL.

Ulu, C., J. E. Smith. 2007. Uncertainty, information acquisition, and technology adoption. Working paper, Duke University.

U.S. Census Bureau. 2006. State estimates by demographic characteristics - age, sex, race, and hispanic origin (SC_EST2005_AGESEX_CIV.csv). http://www.census.gov/popest/datasets.html.

U.S. National Health Interview Survey. 2005. Self-reported influenza vaccination coverage trends 1989-2005 among adults by age group, risk group, race/ethnicity, health-care worker status, and pregnancy status.

Vaccines and Related Biological Products Advisory Committee. 2001-2006. Transcripts and other meeting documents. Retrieved February 21, 2007, http://www.fda.gov/cber/advisory/vrbp/vrbpmain.htm.

Weycker, D., J. Edelsberg, M. Elizabeth Halloran, I. M. Longini, A. Nizam, V. Ciuryla, G. Oster. 2005. Population-wide benefits of routine vaccination of children against influenza. *Vaccine* **23** 1284-1293.

World Health Organization. 2005. Influenza vaccines: WHO position paper. *Weekly Epidemiological Record* **33**(80) 277-288.

World Health Organization. 2007. Influenza. Retrieved February 21, 2007, http://www.who.int/ immunization/topics/influenza/en/index.html.

Wu, J. T., L. M. Wein, A. S. Perelson. 2005. Optimization of influenza vaccine selection. *Operations Research* **53** 456-476.