Manufacturing & Service Operations Management

Contracting for On-Time Delivery in the U.S. Influenza Vaccine Supply Chain

Tinglong Dai, Soo-Haeng Cho, Fuqiang Zhang

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management
science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Contracting for On-Time Delivery in the U.S. Influenza Vaccine Supply Chain

Tinglong Dai
Carey Business School, The Johns Hopkins University, Baltimore, Maryland 21202, dai@jhu.edu

Soo-Haeng Cho
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, soohaeng@andrew.cmu.edu

Fuqiang Zhang
Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130, fzhang22@wustl.edu

Although influenza vaccine shortage is often attributed to low supply, it has been observed that even with abundant supply, a major shortage can still occur because of late delivery. In this paper, motivated by the influenza vaccine industry, we study a supply chain contracting problem in the presence of uncertainties surrounding design, delivery, and demand of the influenza vaccine. In this supply chain, a manufacturer has insufficient incentive to initiate at-risk early production prior to the design freeze because it is a retailer who reaps the most benefits from selling more vaccines delivered on time. Anticipating that late delivery will lead to potential loss in demand, the retailer tends to reduce the order size, which further discourages the manufacturer from making an effort to improve its delivery performance. To break this negative feedback loop, a supply contract needs to achieve two objectives: incentivize at-risk early production and eliminate double marginalization. We find that two commonly observed supply contracts in practice, the delivery-time-dependent quantity flexibility (D-QF) contract and the late-rebate (LR) contract, may fail to coordinate the supply chain because of the tension between these two objectives. To resolve such a tension, we construct a buyback-and-late-rebate (BLR) contract and show that a properly designed BLR contract can not only coordinate the supply chain but also can provide full flexibility of profit division between members of the supply chain. Numerical experiments further demonstrate that the BLR contract significantly improves supply chain efficiency compared to the contracts used in the industry.

Keywords: influenza vaccine; supply contract; on-time delivery; coordination
History: Received: May 13, 2014; accepted: November 17, 2015. Published online in *Articles in Advance* March 30, 2016.

> If you want a [flu] shot, you're gonna have to dance for it.
> —"Dr. Leo Spaceman," *30 Rock*, Season 3, Episode 8

1. Introduction

Behind many conventional products are myriad unconventional challenges. When reflecting on the vaccine industry, James Matthews (2006, p. 19) of Sanofi Pasteur observes, "Even though the seasonal influenza vaccine is considered a conventional vaccine by the industry, new challenges with respect to timing and availability of strains and the composition of the influenza vaccine are the rule." A special feature of the influenza vaccine industry is that a manufacturer does not decide the design of its own product (i.e., the composition of the influenza vaccine). In the United States, for example, the Vaccine and Related Biologic Products Advisory Committee (VRBPAC), in short the Committee hereinafter, which is independent of manufacturers, makes recommendations to the Food and Drug Administration (FDA) about the annual vaccine composition in

February or March of each year for the upcoming flu season that begins the following October. The timing of this decision creates remarkable challenges: On one hand, the production process is complex and highly uncertain; on the other hand, there is a tight time window left between the announcement of the composition and the start of the flu season. These challenges make it extremely difficult to match supply with demand; in particular, supply shortage can occur even when the supply is abundant. As an illustrative example, influenza vaccine coverage recorded a decline to 41% in the 2000–2001 influenza season, compared to 57% in the previous season; meanwhile, 7.5 million vaccine doses, or 10.6% of the total supply, remained unused by the end of the season (Nowalk et al. 2005, O'Mara et al. 2003). Fukuda et al. (2002, p. 235) explain this seemingly paradoxical situation as follows:

> The availability of influenza vaccine [in 2000 and 2001] was significantly lower during [October and November] than in previous years, which left many clinicians and patients unable to find vaccine and led to the

cancellation of many vaccination campaigns. Ironically, in both years, increasing supplies of vaccine became available in December, but the waning levels of demand resulted in substantial surpluses of unused vaccine.

More recently, during the 2014–2015 season, influenza vaccine manufacturers experienced delayed shipments in the United States, resulting in a shortage during the peak demand time in October (Loftus 2014).

A common practice for vaccine manufacturers to improve their delivery performance is to start producing vaccines prior to the Committee's announcement of the vaccine composition (VRBPAC 2002–2014). This option, however, involves the risk that a manufacturer's projected composition may differ from the Committee's decision—in this case, the whole batch of vaccine strains in production will have to be discarded. According to Raymond Fitch, the director of viral manufacturing at Sanofi Pasteur, who spoke in 2009 on behalf of the industry (VRBPAC 2002–2014):

> To make sure that we hit that timing both on distribution timing and total supply, manufacturers have to enter into manufacturing processes in January of that year under a risk condition which is the unselected strains. …Production was initiated by most manufacturers under an at-risk condition because of still pending the strain selection process.

Whereas the manufacturer bears the entire risk associated with this early production, its benefit mostly accrues to a retailer (i.e., a healthcare provider) because the retailer can generate more revenue from vaccines delivered on time by the manufacturer. Thus, a well-designed supply contract needs to provide proper incentives for the manufacturer to improve its delivery performance.

As the initial step of our study, we collected supply contracts between major vaccine manufacturers and two major academic medical centers over the past several years. Most of these contracts were signed in January for vaccines to be delivered for the next flu season starting in October. Our focus in this paper is on this direct distribution channel between manufacturers and retailers, although manufacturers may distribute vaccines through distributors who in turn deliver vaccines to smaller retailers such as physician offices. Table 1 provides representative sample contracts used by two major manufacturers (referred to as A and B, respectively) during three consecutive seasons. For instance, we refer to the contract used by Manufacturer A as the delivery-time-dependent quantity flexibility (D-QF) contract, and the contract used by Manufacturer B during the 2009–2010 season is the so-called late-rebate (LR) contract. It is interesting to observe that the D-QF contract, although new to the literature, resembles the quantity flexibility (QF) contract (see, e.g., Cachon 2003) but differs in that the maximum

Table 1   Sample Contracts in Influenza Vaccine Industry

Manufacturer	Season	Contract terms
A	2010–2011	A proportion of unused doses can be returned for full credit: • doses shipped before October 15: up to 25% of the doses; • doses shipped after October 15: up to 50% of the doses.
	2009–2010	The same as in the 2010–2011 season.
	2008–2009	The same as in the 2010–2011 season except that the cut-off date is November 15.
B	2010–2011	No returns are allowed; no rebate for late-delivered items.
	2009–2010	A 10% rebate is provided for orders shipped after September 30.
	2008–2009	No returns are allowed; no rebate for late-delivered items.

returnable quantity depends on the timing of delivery. We find that manufacturers tend to use their preferred "type" of contracts such as D-QF and LR contracts, although some specific prices or terms were slightly different, possibly as a result of negotiations.

There are a couple of observations from Table 1. First, the two manufacturers used different types of contracts. Second, even for the same manufacturer, contract types varied across years. It appears that the industry has been experimenting with different types of contracts. In light of these observations, one may raise several questions about how to manage this supply chain: How does the inclusion of multiple types of uncertainties and the production timing decision complicate the incentive alignment problem in the supply chain? Do those contracts used in practice perform well? For instance, the D-QF contract adopts a delivery-time dependent term to encourage early production by the manufacturer. Does this added complexity in the contract design lead to coordination of the supply chain? What contracting options should be recommended to improve supply chain efficiency and social welfare?

To answer these questions, we develop an analytical model that captures the following three key sources of uncertainties in this supply chain: (1) The product *design* is exogenous to a manufacturer because the Committee determines the composition of the influenza vaccine. Thus, if the manufacturer begins its production prior to a design freeze, then it faces the risk associated with product design. (2) The *delivery* lead time required for manufacturing and distributing vaccines is long (usually six to eight months) and uncertain. Because of the complex processes of production, testing, releasing, and distribution, a manufacturer has to make its production decision way in advance of the demand season, but its delivery of vaccine can still be delayed, especially when it begins its production after the design

uncertainty is resolved. (3) The *demand* is time sensitive and uncertain. The Centers for Disease Control and Prevention (CDC) notes the time sensitivity of the demand as follows (CDC 2014): "Manufacturers with vaccine coming off the production line in middle or late November or later may not be able to sell it all and providers receiving vaccine in this same time frame may not be able to convince patients to receive it." When a retailer signs a contract, typically in January, it faces an uncertain demand for the next flu season that starts the following October.

Under a myriad of these uncertainties, we find that various well-studied contracts are not effective in inducing satisfactory delivery performance because of a *negative feedback loop* in the firms' incentives: Since the benefit of on-time delivery mostly accrues to the retailer, the manufacturer lacks the motivation to improve the on-time delivery performance, which leads to potential loss in demand and induces the retailer to order a low quantity. We then analyze the D-QF and LR contracts used in the industry that have special penalty clauses for late delivery. Although each penalty clause provides intuitive incentives for early production, we find that both contracts may fail to coordinate the supply chain under realistic settings. This is because of the inherent tension between the two objectives to achieve coordination: incentivizing the manufacturer to undertake at-risk early production and inducing the retailer to choose adequate order quantity. Given the interdependence of the two decisions, one needs to carefully orchestrate incentives for both parties to maximize supply chain efficiency. Finally, we propose a new contract type, called the buyback-and-late-rebate (BLR) contract that builds on the strengths of the D-QF and LR contracts. This contract is able to coordinate the supply chain; meanwhile, it also allows a flexible division of payoff between the two parties. Our numerical analysis suggests that the use of a coordinating contract improves the profitability of the supply chain by 12.14% and 15.55% on average over the LR and wholesale price contracts, respectively, as used in this industry.

2. Literature Review

This work draws on and contributes to the following two streams of literature. First, we contribute to the rich literature of supply contracts by evaluating various (well-known and new) types of contracts in the new environment where a supply chain faces uncertainties in design, delivery, and demand. Second, we contribute to the influenza vaccine operations literature by studying a contracting problem between an influenza vaccine manufacturer and a healthcare provider based on the real contracts used in practice.

Supply contracts have been studied extensively. Below we review only the papers that are most related

to this paper, while referring readers to Cachon (2003) for a comprehensive review of early work. Our paper is related to the papers that study buyback/returns, quantity flexibility contract, and rebates, including Arya and Mittendorf (2004), Lariviere (1998), Padmanabhan and Png (1997), Pasternack (1985), Taylor (2002), and Tsay (1999). Whereas these papers focus on *demand uncertainty*, our paper also addresses uncertainties in delivery timing and product design.

The issue of *on-time delivery* has been studied from various angles. Grout and Christy (1993) study purchasing contracts in a just-in-time setting where the delivery performance is controlled by a supplier and show that, under delivery uncertainty only, a bonus scheme improves on-time delivery. Cachon and Zhang (2006) consider the sourcing problem of a buyer whose operating costs are affected by both procurement price and delivery lead time, and characterize the optimal procurement mechanisms and two simple but effective strategies. Hwang et al. (2014) consider the per-unit penalty contract adopted by retailers that resembles the late-rebate (LR) contract studied in our paper, but in their model the supplier has a single production mode and does not face demand and delivery uncertainties. Our paper differs significantly from and thus enriches this literature by studying a delivery-design trade-off in the context of the vaccine industry.

To mitigate delivery risk, a manufacturer in our model operates in dual production modes. This resembles the setting of Donohue (2000) that models a fast fashion supply chain with two production modes: one is cheap but has a long lead time, and another is expensive but more responsive to market demand. Three key differences separate our paper from Donohue's. First, in our paper, early production helps improve delivery performance, whereas in Donohue's case it reduces the production cost. Second, we consider uncertainties in design, delivery, and demand, whereas Donohue considers only demand uncertainty. Third, our paper analyzes various supply contracts observed in the vaccine industry, whereas Donohue focuses on the wholesale price contract.

The second stream of related literature studies various operational issues in the influenza vaccine supply chain. Chick et al. (2008) show that if a central government can select a fraction of a population to vaccinate, then the government can use a cost-sharing contract to induce a manufacturer to produce the socially optimal quantity. Deo and Corbett (2009) analyze the effect of yield uncertainty on competition among vaccine manufacturers. Cho (2010) studies the Committee's problem of choosing an optimal vaccine composition with dynamic information updating. Arifoğlu et al. (2012) study the impact of yield uncertainty and self-interested consumers on the inefficiency in the supply chain, and they analyze the effectiveness of government

interventions through partial centralization. Adida et al. ([2013](#)) and Mamani et al. ([2012](#)) study how the government can induce a socially optimal vaccine coverage through subsidies to a manufacturer and consumers. Chick et al. ([2012](#)) extend Chick et al. ([2008](#)) by considering a setting where the manufacturer has to satisfy the exact demand determined by the government: If a low production yield leads to a shortfall, then the manufacturer is required to make up the difference. They show that the supply chain is coordinated when the government's additional administrative expense is transferred to the manufacturer.

Our paper makes the following contributions to this literature. First, our paper is the first to consider a healthcare provider in the U.S. influenza vaccine supply chain who places an order to a manufacturer and then distributes vaccines to consumers. This research perspective is shared by the case study by Deo et al. ([2012](#)) on vaccine purchasing at NorthShore University HealthSystem. Based on the *real* contracts used in practice, we analyze various contracts between a healthcare provider and a manufacturer. Second, we employ several new modeling elements supported by industry evidence such as uncertain delivery timing, early production mode associated with design risks, and time-sensitive uncertain demand. Although the literature often attributes vaccine shortage to low supply, our model reflects the observation that even abundant supply may still result in shortage because of late delivery. Finally, whereas the previous literature studies the effectiveness of *potential* government interventions through partial centralization or subsidies, we shed light on improvement opportunities through coordinating contracts between firms in this supply chain.

3. Modeling Framework

We consider a supply chain consisting of two risk-neutral firms, a manufacturer and a retailer. As commonly assumed in the literature (e.g., Arifoğlu et al. [2012](#), Chick et al. [2008](#)), the retailer sells a product at a fixed price p. The associated demand, denoted by ξ, follows a distribution $F(\cdot)$ with a density $f(\cdot)$. To model the time sensitivity of the demand, we consider two selling periods: an *ideal* period and a *late* period. Demand arrives during the ideal period, but if there is unmet demand by the end of the ideal period because of inadequate supply, then a proportion $\gamma \in (0, 1)$ of the unserved customers will not return for vaccination, whereas the rest will return during the late period. Thus, the parameter γ captures the time sensitivity of demand. Based on these demand characteristics, the retailer determines the order quantity Q, which incurs an administrative cost of $c_o(\geq 0)$ per unit that captures the burden of processing the order.

The manufacturer operates in dual production modes: "regular" and "early" with respective subscripts r and e. Under the *regular* production mode, the manufacturer has an uncertain delivery lead time and cannot always deliver the product in a timely fashion. With probability $\alpha \in (0, 1)$, the delivery is on time (i.e., satisfying the demand during the ideal period); with probability $(1 - \alpha)$, the delivery is late (i.e., satisfying the demand during the late period). The manufacturer also has the *early* production mode, in which the manufacturer starts at-risk production before the product design (i.e., vaccine composition) is finalized. The early production mode guarantees on-time delivery but is vulnerable to design uncertainty: With probability $\beta \in (0, 1)$, the early production uses the same composition as the finalized one, and with probability $(1 - \beta)$, the early production uses a different composition, in which case the whole batch of vaccine in production has to be discarded. The respective unit production costs under the regular and early production modes are c_r and c_e. The manufacturer first decides an early production quantity, denoted by Q_e. Given Q and Q_e, a regular production quantity, denoted by Q_r, can be one of the following two quantities, depending on the outcome of early production:

$$Q_r = \begin{cases} Q - Q_e & \text{with probability } \beta \\ Q & \text{with probability } (1 - \beta). \end{cases} \quad (1)$$

Below we specify the sequence of events (see Figure [1](#)):

• $t = 1$: The retailer determines an order quantity Q, and then the manufacturer determines an early production quantity Q_e.

• $t = 2$: Upon the release of the final product design, the manufacturer determines a regular production quantity Q_r according to Equation ([1](#)).

• $t = 3$ (the ideal period): With probability β, early-production outputs are delivered to the retailer during this period. In addition, with probability α, regular-production outputs are delivered during this period.

• $t = 4$ (the late period): Among the unserved customers in the ideal period, a fraction $(1 - \gamma)$ of them will return to seek vaccination. On the supply side, with probability $(1 - \alpha)$, regular production outputs are delivered to satisfy such residual demand.

For notational convenience, we define an indicator $I \in \{0, 1\}$ as

$$I = \begin{cases} 1 & \text{if products from the regular production} \\ & \quad \text{are delivered on time,} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we define another indicator $J \in \{0, 1\}$ to represent whether or not the early production matches the Committee's recommendation. We have $\Pr(I = 1) = \alpha$, $\Pr(I = 0) = 1 - \alpha$, $\Pr(J = 1) = \beta$, and $\Pr(J = 0) = 1 - \beta$.

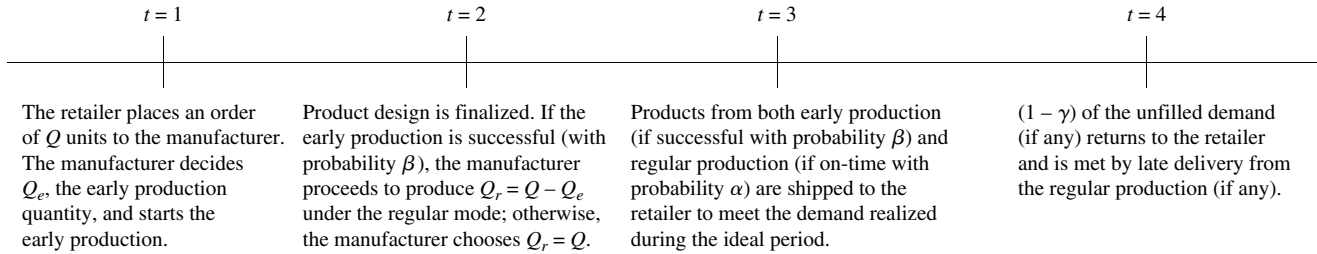**Figure 1   Sequence of Events Under the Dual Production Modes**

$t = 1$	$t = 2$	$t = 3$	$t = 4$
The retailer places an order of Q units to the manufacturer. The manufacturer decides Q_e, the early production quantity, and starts the early production.	Product design is finalized. If the early production is successful (with probability β), the manufacturer proceeds to produce $Q_r = Q - Q_e$ under the regular mode; otherwise, the manufacturer chooses $Q_r = Q$.	Products from both early production (if successful with probability β) and regular production (if on-time with probability α) are shipped to the retailer to meet the demand realized during the ideal period.	$(1 - \gamma)$ of the unfilled demand (if any) returns to the retailer and is met by late delivery from the regular production (if any).

Table 2 lists three cases depending on the realization of I and J.

Before proceeding to our analysis, we offer a few remarks on the model. First, we abstract away from modeling the detailed epidemiology and consumers' vaccination decisions that constitute the demand-forming process. This is for both analytical tractability and practical relevance: It is well known that it is extremely difficult to forecast flu activities in the next flu season starting in October based on epidemic data until January (e.g., Lofgren et al. 2007), and the practitioners we interviewed placed orders in January simply using forecasts from the previous demand data. Second, our qualitative results will not change if the value of γ depends on the quantity delivered on time, reflecting the observation that some consumers' vaccine-seeking behavior may depend on the vaccination and thus infection conditions of the population in the ideal period (see Arifoğlu et al. 2012). Third, we follow the convention of the supply chain coordination literature (e.g., Cachon and Lariviere 2001, Cachon 2003) by adopting the forced compliance regime under which the manufacturer must supply Q as ordered by the retailer and by assuming that both firms are risk-neutral.

As a benchmark, we first analyze the first-best scenario in which a central decision maker jointly determines the order quantity Q and the early production quantity Q_e to maximize the supply chain profit. Let $\pi_S(Q, Q_e)$ denote the expected profit of the supply chain and $Z(Q, Q_e)$ the ex post sales quantity. We refer to the solution of the following problem, denoted by (Q^{FB}, Q_e^{FB}), as the first-best solution:

$$\underset{Q > 0,\, 0 \le Q_e \le Q}{\text{maximize}} \ \pi_S(Q, Q_e)$$
$$= pE[Z(Q, Q_e)] - [(c_r + c_o)Q + (c_e - \beta c_r)Q_e]. \quad (2)$$

Table 2   Three Cases of Delivery and Early Production Outcome

Case 1.	$I = 1$ (on-time delivery)	With probability α
Case 2.	$I = 0$ (late delivery) and $J = 1$ (successful early production)	With probability $(1 - \alpha)\beta$
Case 3.	$I = 0$ (late delivery) and $J = 0$ (unsuccessful early production)	With probability $(1 - \alpha)(1 - \beta)$

In (2), $[(c_r + c_o)Q + (c_e - \beta c_r)Q_e]$ is the total expected cost, which is derived from

$$c_e Q_e + c_r E[Q_r] + c_o Q$$
$$= c_e Q_e + c_r [\underbrace{\beta(Q - Q_e)}_{J=1} + \underbrace{(1 - \beta)Q}_{J=0}] + c_o Q.$$

We assume that $c_e > \beta c_r$—which includes the case that $c_e = c_r$—to focus on a realistic situation where there is a trade-off between the delivery advantage of early production and the informational advantage of regular production. This assumption guarantees the following result: $Q_e^{FB} < Q^{FB}$; see Lemma A1 in Online Appendix A (available as supplemental material at http://dx.doi.org/10.1287/msom.2015.0574). Furthermore, the first-best, early-production quantity Q_e^{FB} is positive only when the delivery from the regular production mode is sufficiently unreliable, the early production mode is not very expensive, and the design uncertainty is not overly high. We focus on the interesting case where $Q_e^{FB} > 0$ because $Q_e^{FB} = 0$ means that the early production is not a viable option (in contrast to the common industry practice).

Next, we analyze a decentralized supply chain under different supply contracts. A contract is said to *coordinate* the supply chain if it induces the first-best solution (Q^{FB}, Q_e^{FB}) from the firms comprising the supply chain. One major performance metric we use in evaluating supply contracts is the efficiency of the supply chain, which is defined as the ratio of the supply chain's expected profit under a contract to that in the first-best scenario.

3.1.   Preliminary: Wholesale Price Contract

To understand key challenges behind the coordination of the vaccine supply chain, in this section, we first analyze the wholesale price contract that was used by Manufacturer B during the 2008–2009 and 2010–2011 seasons (see Table 1). With a wholesale price w, the manufacturer's profit is $\pi_M^W(Q, Q_e) = wQ - c_r E[Q_r] - c_e Q_e = (w - c_r)Q - (c_e - \beta c_r)Q_e$, and the retailer's profit is $\pi_R^W(Q, Q_e) = pE[Z(Q, Q_e)] - (w + c_o)Q$.

As compared with a typical supply chain with only demand uncertainty, a contract requires the coordination of the manufacturer's production decision as well as the retailer's ordering decision. The following

lemma shows that these two critical decisions are interdependent:

LEMMA 1.
(i) $\partial E[Z(Q, Q_e)]/\partial Q_e \geq 0$, $\partial E[Z(Q, Q_e)]/\partial Q > 0$.
(ii) $\partial^2 E[Z(Q, Q_e)]/\partial Q \partial Q_e > 0$.

Lemma 1(i) means that the expected sales quantity increases both in the manufacturer's early production quantity Q_e and in the retailer's order quantity Q. Lemma 1(ii), however, is less intuitive. It states that the marginal gain in the expected sales quantity from a higher order quantity Q increases in the early production quantity Q_e, because the retailer can expect more on-time delivered products from increased Q_e and hence a lower chance of lost sales. Interestingly, Lemma 1(ii) also suggests that the marginal gain from early production increases in the order quantity Q as well. In other words, the decisions Q_e and Q are complementary to each other: A higher order quantity Q will make early production more beneficial to the supply chain and vice versa. Although this might appear as a positive side of the supply chain, it actually reveals a negative feedback loop in the firms' incentives: On one hand, when the manufacturer bears the risk associated with early production, it lacks the incentive to improve the delivery performance, which leads to potential loss in demand; on the other hand, the demand loss induces the retailer to reduce its order quantity, which further discourages the manufacturer from making an effort to improve on-time delivery.

Because of the negative feedback loop, it is straightforward to show that a wholesale price contract (used by Manufacturer B) cannot coordinate the supply chain. Similarly, the conventional contracts such as buyback, revenue sharing, quantity discount, and sales rebate do not coordinate the supply chain, either. Furthermore, two-parameter contracts may perform even worse than a wholesale price contract (e.g., for any revenue-sharing contract with a positive revenue share, there always exists a wholesale price contract that can achieve a higher supply chain efficiency), which counters intuition that adding one more parameter to the supply contract would lead to higher supply chain efficiency. One intuitive way to overcome the delivery challenge in this supply chain is to impose a penalty on late delivery. Practitioners surely understand the importance of motivating early production, and as discussed in §1, the D-QF and LR contracts that are used in the industry include such a penalty in two different forms (i.e., buyback and rebate, respectively). Are those late penalty terms sufficient to induce the firms to make decisions to optimize supply chain efficiency? We examine this question next in §4. After analyzing these two contracts used in practice, we propose and analyze the BLR contract in §5. Then we evaluate the performance of different contracts in §6. Finally, in §7 we consider various extensions of the main model.

4. Analysis of D-QF and LR Contracts

In this section we analyze the D-QF and LR contracts, focusing on whether these contracts facilitate the coordination of the supply chain, and whether they allow for flexible division of total supply chain profits between the firms. We assume that the manufacturer's early production quantity is not verifiable by the retailer and hence cannot be contracted on.

4.1. Delivery-Time-Dependent Quantity Flexibility (D-QF) Contract

Under a D-QF contract, the retailer is allowed to return its leftover inventory at full price up to some level, referred to as return allowance, that depends on the timing of delivery. Although this contract has been adopted in the U.S. influenza vaccine supply chain (e.g., Manufacturer A in Table 1), it has not been reported in the literature.

Let Y_1 be the quantity delivered by the end of the ideal period, and Y_2 the quantity delivered after the ideal period. The return allowance is then equal to $\kappa_1 Y_1 + \kappa_2 Y_2$, where $\kappa_1, \kappa_2 \in [0, 1]$ are the returnable proportions of the quantities delivered on time and late, respectively. Hence, the return allowance can be represented as

$$\kappa_1 Y_1 + \kappa_2 Y_2 = \begin{cases} \kappa_1 Q & \text{(Case 1)} \\ \kappa_1 Q_e + \kappa_2 (Q - Q_e) & \text{(Case 2)} \\ \kappa_2 Q & \text{(Case 3).} \end{cases}$$

When $\kappa_1 = \kappa_2$, the D-QF contract is reduced to the quantity flexibility (QF) contract previously studied in the literature. We denote by $R_d(Q, Q_e)$ the total returning quantity at the end of a demand season under the D-QF contract: $R_d(Q, Q_e) = \min\{\kappa_1 Y_1 + \kappa_2 Y_2, Q - Z(Q, Q_e)\}$; see Online Appendix A for its detailed characterization. The transfer payment from the manufacturer to the retailer is $T_d(Q, Q_e) = w \cdot R_d(Q, Q_e)$, where w is the wholesale price.

Now we characterize the parameters of a D-QF contract that coordinate the supply chain. For ease of exposition, let us define the following three numbers:

$$\xi^{(1)} \equiv (1 - \kappa_2)Q + (\kappa_2 - \kappa_1)Q_e \text{ so that}$$
$$\xi \geq \xi^{(1)} \Leftrightarrow Q - \xi \leq \kappa_1 Q_e + \kappa_2 (Q - Q_e);$$

$$\xi^{(2)} \equiv [(1 - \kappa_2)Q + (\kappa_2 - \kappa_1 - \gamma)Q_e]/(1 - \gamma) \text{ so that}$$
$$\xi \geq \xi^{(2)} \Leftrightarrow Q - \{Q_e + (1 - \gamma)(\xi - Q_e)\}$$
$$= Q - \gamma Q_e - (1 - \gamma)\xi \leq \kappa_1 Q_e + \kappa_2 (Q - Q_e);$$

$$\xi^{(3)} \equiv (Q - \gamma Q_e)/(1 - \gamma) \text{ so that}$$
$$\xi \geq \xi^{(3)} \Leftrightarrow Q - Q_e \leq (1 - \gamma)(\xi - Q_e);$$

where $\xi^{(1)}$ is the demand level at which the leftover inventory is equal to the return allowance when $\xi < Q_e$, $\xi^{(2)}$ is the demand level at which the leftover inventory is equal to the return allowance when $\xi \geq Q_e$, and $\xi^{(3)}$ is the demand level above which there will be no leftover inventory (implying that $\xi^{(2)} < \xi^{(3)}$).

PROPOSITION 1.

(i) *Any D-QF contract with $\kappa_2 \le \kappa_1$ (where $\kappa_1, \kappa_2 \in [0,1]$) cannot coordinate the supply chain.*

(ii) *Suppose both firms' objective functions are unimodal. Then a D-QF contract coordinates the supply chain if and only if one of the following two conditions is satisfied:*
(A) $\kappa_1 Q_e^{FB} + \kappa_2 (Q^{FB} - Q_e^{FB}) < Q^{FB} - Q_e^{FB}$, and at $(Q, Q_e) = (Q^{FB}, Q_e^{FB})$,

$(A_M):$ $\quad (1-\alpha)\beta[(\kappa_2 - \kappa_1 - \gamma)F(\xi^{(2)}) + \gamma F(Q_e)]$
$$= \frac{c_e - \beta c_r}{w},$$
$(A_R):$ $\quad \alpha[F(Q) - (1-\kappa_1)F((1-\kappa_1)Q)]$
$$+ (1-\alpha)\beta[F(\xi^{(3)}) - (1-\kappa_2)F(\xi^{(2)})]$$
$$+ (1-\alpha)(1-\beta)\left[F\left(\frac{Q}{1-\gamma}\right)\right.$$
$$\left. - (1-\kappa_2)F\left(\frac{(1-\kappa_2)Q}{1-\gamma}\right)\right] = 1 - \frac{c_r}{w}.$$

(B) $\kappa_1 Q_e^{FB} + \kappa_2 (Q^{FB} - Q_e^{FB}) \ge Q^{FB} - Q_e^{FB}$, *and at* $(Q, Q_e) = (Q^{FB}, Q_e^{FB})$,

$(B_M):$ $\quad (1-\alpha)\beta(\kappa_2 - \kappa_1)F(\xi^{(1)}) = \frac{c_e - \beta c_r}{w},$
$(B_R):$ $\quad \alpha[F(Q) - (1-\kappa_1)F((1-\kappa_1)Q)]$
$$+ (1-\alpha)\beta[F(\xi^{(3)}) - (1-\kappa_2)F(\xi^{(1)})]$$
$$+ (1-\alpha)(1-\beta)\left[F\left(\frac{Q}{1-\gamma}\right)\right.$$
$$\left. - (1-\kappa_2)F\left(\frac{(1-\kappa_2)Q}{1-\gamma}\right)\right] = 1 - \frac{c_r}{w}.$$

Part (i) suggests that a necessary condition for a D-QF contract with (κ_1, κ_2) to coordinate the supply chain is $0 \le \kappa_1 < \kappa_2 \le 1$. The condition looks intuitive at first glance, but it merits some discussion. It first implies that $\kappa_2 > 0$, meaning that return allowance for late-delivered items must be positive. This is required for the manufacturer to bear at least some level of delivery risk. Second, the condition requires $\kappa_2 > \kappa_1$, meaning that return allowance for late delivery must be higher than that for on-time delivery. As discussed earlier in this section, the D-QF contract with $\kappa_2 = \kappa_1 (> 0)$ is reduced to the QF contract. Thus, this result implies that the QF contract can never coordinate the supply chain. It is worth noting that the QF contract does provide the manufacturer with some degree of incentive to undertake at-risk early production because of the time-sensitive demand; i.e., even with $\kappa_2 = \kappa_1$, the manufacturer is better off with on-time delivery because on-time delivered vaccines can serve the demand in both ideal and late periods, whereas late-delivered vaccines can serve only the lower demand in the late period, resulting in higher returns to the manufacturer.

Therefore, this condition suggests that a D-QF contract needs to provide a substantial level of incentive to motivate the manufacturer to undertake at-risk early production. This is consistent with the practice: in Table 1, Manufacturer A uses $\kappa_2 = 50\% > \kappa_1 = 25\%$.

Part (ii) further presents necessary and sufficient conditions for a D-QF contract to coordinate the supply chain. Conditions (A) and (B) correspond to the scenario under which the total return allowance is relatively low and high, respectively. This can be seen by noting that $\kappa_1 Q_e^{FB} + \kappa_2 (Q^{FB} - Q_e^{FB})$ and $Q^{FB} - Q_e^{FB}$ are the return allowance and the late-delivered quantity for Case 2, respectively, when Q^{FB} and Q_e^{FB} are chosen. These conditions illustrate two objectives to achieve for channel coordination: mitigate the manufacturer's risk associated with early production and thus induce the manufacturer to improve its on-time delivery performance (as reflected in (A_M) and (B_M)), and overcome double marginalization and thus induce the retailer to order more (as reflected in (A_R) and (B_R)). Observe that (A_M) and (B_M) contain only the probability for Case 2, $(1-\alpha)\beta$, because early production helps improve on-time delivery only in the case when regular production yields late delivery and early production is successful. In contrast, (A_R) and (B_R) contain the probability for each of Cases 1–3 (i.e., α for Case 1, $(1-\alpha)\beta$ for Case 2, and $(1-\alpha)(1-\beta)$ for Case 3; see Table 2) because the order quantity Q affects sales and inventory outcomes in all three cases.

Conditions (A) and (B) can be interpreted as follows. On the manufacturer's side, the contract needs to motivate the manufacturer to undertake at-risk early production to improve on-time delivery performance. To this end, $(\kappa_2 - \kappa_1)$ needs to be sufficiently high (given κ_2) to satisfy (A_M) and (B_M). On the retailer's side, the contract needs to eliminate double marginalization. A D-QF contract lessens double marginalization by allowing the retailer to return leftover up to the total return allowance, $\kappa_1 Y_1 + \kappa_2 Y_2$. Thus, κ_1 and κ_2 must be sufficiently high to incentivize the retailer to increase its order size. Especially, when on-time delivered units Y_1 is expected to be larger than late delivered units Y_2 (see Online Appendix A), a D-QF contract must specify return allowance κ_1 sufficiently high to be effective in addressing double marginalization. This can be seen from the first term in the left-hand side of (A_R) or (B_R), which increases in κ_1.

Finally, we note that a tension may arise between the two objectives for channel coordination. The manufacturer's incentive to improve delivery performance increases with $(\kappa_2 - \kappa_1)$, but the retailer places a large order especially when κ_1 is high. As a result, there may not exist a coordinating D-QF contract. Such situations may arise in the following circumstances when the value of early production to the manufacturer is low: (i) regular production is highly reliable (i.e., α is high);

Figure 2   (Color online) Existence or Nonexistence of Coordinating D-QF Contracts
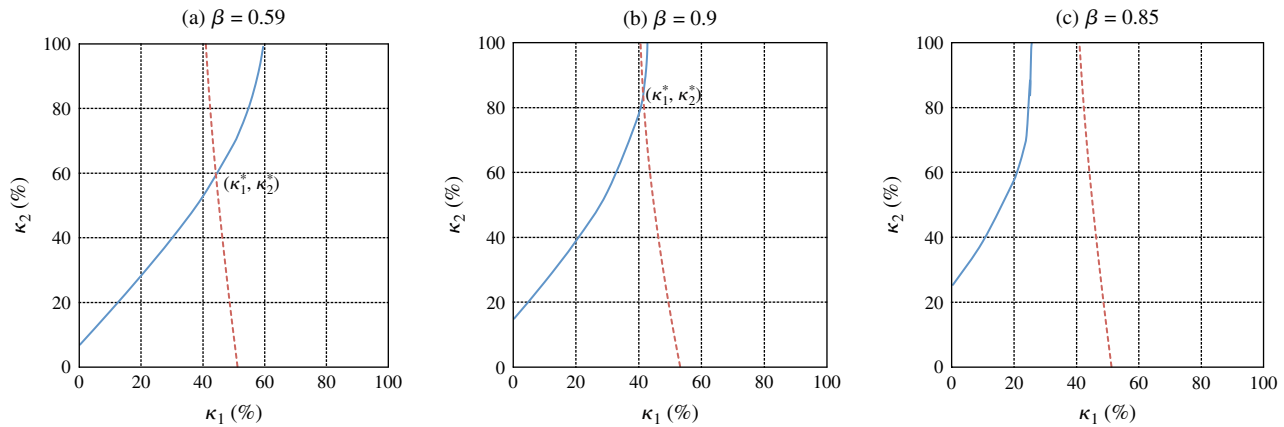
Notes. Parameters used: $c_r = c_e = \$3$, $c_o = \$0$, $w = \$8$, $p = \$18$, $\alpha = 0.8$, $\gamma = 0.3$. The demand follows a gamma distribution with a density of $f(\xi) = (\beta_g)^{-\alpha_g} e^{-\xi/\beta_g} \xi^{\alpha_g - 1}/\Gamma(\alpha_g)$, where $\alpha_g = 4$ and $\beta_g = 0.25$. A solid curve represents D-QF contracts that satisfy the first objective (i.e., induce the manufacturer to choose Q_e^{FB}), and a dotted curve represents D-QF contracts that satisfy the second objective (i.e., induce the retailer to order Q^{FB}).

(ii) final product design is difficult to predict (i.e., β is low); (iii) demand is not so sensitive to time (i.e., γ is low); or (iv) cost difference between the two production modes is high (i.e., $c_e - \beta c_r$ is high). This can be seen in conditions (A) and (B); for example, when β is low, ceteris paribus, a high $(\kappa_2 - \kappa_1)$ would be necessary to satisfy (A_M) and (B_M), but a low β makes it difficult to satisfy (A_R) or (B_R). Figure 2 illustrates how the existence of a coordinating D-QF contract and its contract parameters (κ_1^*, κ_2^*) change with β. Coordinating contract parameters (κ_1^*, κ_2^*) are obtained in Figure 2(a)–(b) where the solid and dotted curves cross, whereas there exists no coordinating D-QF contract in Figure 2(c) with no crossing point.

REMARK 1. Part (ii) of Proposition 1 builds on the unimodality of the firms' objective functions. The retailer's objective function is unimodal under an arbitrary probability distribution. We can analytically show that under many distributions, a sufficient condition for the manufacturer's objective function to be unimodal is that $\kappa_2 - \kappa_1$ is not too large; for example, under uniform distribution, a sufficient condition is that $\kappa_2 - \kappa_1 \leq \min\{\sqrt{\gamma/(1-\gamma)}, \sqrt{2\gamma}\}$. We can numerically demonstrate that the manufacturer's objective function is unimodal for a wide range of scenarios.

Flexibility of Profit Division. We now examine the issue of profit division between the two firms under coordinating D-QF contracts. We show in the following that any profit division is possible by choosing a wholesale price w. Under a D-QF contract, the manufacturer's expected profit is $\pi_M^{D-QF} = wQ - c_r Q - (c_e - \beta c_r)Q_e - wE[R_d(Q, Q_e)]$, and hence the manufacturer's profit share ϕ_M is given by

$\phi_M(w)$
$$= \frac{wQ^{FB} - wE[R_d(Q^{FB}, Q_e^{FB})] - (c_r + c_o)Q^{FB} - (c_e - \beta c_r)Q_e^{FB}}{pE[Z(Q^{FB}, Q_e^{FB})] - (c_r + c_o)Q^{FB} - (c_e - \beta c_r)Q_e^{FB}}.$$

As the wholesale price w approaches the retail price p, $\phi_M(w)$ approaches one because (i) the expected returning quantity $E[R_d(Q, Q_e)]$ cannot exceed the total leftover inventory, so $E[R_d(Q, Q_e)] \leq Q - E[Z(Q, Q_e)]$, which can be rewritten as $Q - E[R_d(Q, Q_e)] \geq E[Z(Q, Q_e)]$; and (ii) $wQ^{FB} - wE[R_d(Q^{FB}, Q_e^{FB})] = w\{Q^{FB} - E[R_d(Q^{FB}, Q_e^{FB})]\} \to p\{Q^{FB} - E[R_d(Q^{FB}, Q_e^{FB})]\} \geq pE[Z(Q^{FB}, Q_e^{FB})]$. On the other hand, as w approaches $c_r + c_o + (c_e - \beta c_r)Q_e^{FB}/Q^{FB}$, $\phi_M(w)$ approaches zero because π_M^{D-QF} approaches zero. Given that profit functions are continuous in w, clearly any profit allocation is possible by choosing w.

Our numerical analysis further shows that if a coordinating D-QF contract exists for a given set of parameters, then there is a wide range of wholesale prices to choose from. For example, for the parameters used in Figure 2(a), a D-QF contract is capable of coordinating the supply chain for any w between \$3.18 and \$18, implying that a change of w does not intensify the tension between the two contracting objectives. To see this, observe from the right-hand sides of (A_M) and (B_M) in Proposition 1 that as w decreases, $\kappa_2 - \kappa_1$ has to be higher. But the right-hand sides of (A_R) and (B_R) suggest that as w decreases, κ_1 needs to be lower.

4.2. Late Rebate (LR) Contract
In the influenza vaccine industry, we observe the usage of an LR contract under which there is a rebate for orders shipped after the ideal vaccination period. For example, Manufacturer B provided a 10% rebate for late-delivered items during the 2009–2010 season (see Table 1). An LR contract is simple to implement because it solely relies on the quantity of late-delivered products.

We denote by ρ_{LR} a proportion of the wholesale price w that the manufacturer rebates to the retailer for late-delivered items. Under this contract, the manufacturer's

expected profit and the retailer's expected profit can be expressed as follows:

$$
\begin{aligned}
\pi_M^{LR} ={}& (w - c_r)Q - (c_e - \beta c_r)Q_e \\
& - \rho_{LR} w(1 - \alpha)(Q - \beta Q_e); \\
\pi_R^{LR} ={}& p \cdot E[Z(Q, Q_e)] - [w + c_o - \rho_{LR} w(1 - \alpha)]Q \\
& - \rho_{LR} w \beta(1 - \alpha)Q_e.
\end{aligned}
\tag{3}
$$

In general, we can show that there does not exist a coordinating LR contract under an arbitrarily chosen w. This is because under an LR contract, the manufacturer would benefit from a decreased marginal cost associated with early production. However, the manufacturer's objective function under an LR contract does not contain $E[Z(Q, Q_e)]$ (expected selling quantity), meaning that it does not directly benefit from the sales of flu vaccine. Formally, we have the following result about the LR contract.

PROPOSITION 2. *The LR contract does not coordinate the supply chain unless the wholesale price $w = c_e/\beta$ and the rebate level $\rho_{LR}^* = (c_e - \beta c_r)/[(1 - \alpha)c_e] \in (0, 1]$, in which case the retailer fully bears the risk associated with design uncertainty and takes all the supply chain profit.*

Proposition 2 shows that there exists a coordinating LR contract only when the retailer has a dominating bargaining power and earns all supply chain profit. This contract requires a wholesale price to be above the unit production cost, since $w = c_e/\beta$ is greater than both c_e and c_r because $\beta < 1$ and $c_e > \beta c_r$ (see §3). Furthermore, it requires $\rho_{LR} = \rho_{LR}^*$ to induce the manufacturer to choose $Q_e = Q_e^{FB}$ because in this case its expected unit cost of early production ($c_e - \beta c_r$) is fully offset by its expected unit benefit from early production $(1 - \alpha)\beta \rho_{LR}^* w$ as shown in (3). In other words, the LR contract effectively transfers all the risk from early production to the retailer. Lastly, note that (1) $\rho_{LR}^* > 0$ is a result of the assumption $c_e > \beta c_r$ (see §3); (2) $\rho_{LR}^* \le 1$ when $c_e/\beta \le c_r/\alpha$, meaning that to achieve on-time delivery, the expected unit cost of early production (c_e/β) is lower than that of regular production (c_r/α). In other words, early production is more cost efficient than regular production in achieving on-time delivery. When this condition is violated, $\rho_{LR}^* > 1$, and hence LR contracts cannot achieve the first-best outcome.

In reality, a retailer rarely has a dominating bargaining power against a large manufacturer, so LR contracts are unlikely to achieve the first-best outcome. An LR contract simply penalizes the manufacturer for late delivery to incentivize the manufacturer to undertake at-risk early production, but unless the wholesale price is very low, it lacks the ability to achieve the second objective of overcoming double marginalization, thus failing to induce the retailer to place a large order.

By Proposition 2, an LR contract would perform well when a retailer has a strong bargaining power such that the wholesale price is set close to c_e/β. In addition, it is intuitive that the retailer's lack of incentive to place a large order is less severe when facing lower demand uncertainty. We can formally show that under low demand uncertainty, a properly designed LR contract can nearly coordinate the supply chain.

5. Buyback-and-Late-Rebate (BLR) Contract

So far we have analyzed three contracts most commonly used in practice, namely, wholesale price, D-QF, and LR contracts. In view of the two key objectives in contract design (eliminating double marginalization and incentivizing the manufacturer's at-risk early production), none of these contracts is satisfactory because they may not achieve the first-best outcome because of the tension between these two objectives. Thus, we investigate a combination of contract structures that have proven to be practical to implement in the influenza vaccine industry (i.e., D-QF and LR contracts).

Using this approach, we now propose a buyback-and-late-rebate (BLR) contract that combines a late rebate term with a buyback term. Under the BLR contract, whereas the late rebate component is the same as the LR contract, the buyback component is a variant of the D-QF contract in that it sets $\kappa_1 = \kappa_2 = 100\%$ and provides a partial buyback credit $b < w$ per unit. Thus, the BLR contract offers partial-credit full-quantity buyback of leftover inventory and a rebate for late-delivered vaccines. Our interviews with vaccine contracting practitioners confirm that the BLR contract can be easily implemented because (1) the quantity of late deliveries is embedded in healthcare providers' purchasing and accounting systems, and (2) vaccine manufacturers arrange trucks to pick up unused leftover inventory from healthcare providers at the end of each season.

We use $\rho \in (0, 1)$ such that $\rho \cdot w$ is the rebate from the wholesale price w for a late-delivered unit. Thus, the expected transfer payment from the manufacturer to the retailer is represented as

$$
\begin{aligned}
& T_{BLR}(Q, Q_e; b, \rho) \\
& = b \cdot \{Q - E[Z(Q, Q_e)]\} \\
& \quad + \rho w[(1 - \alpha)\beta(Q - Q_e) + (1 - \alpha)(1 - \beta)Q]. \quad (4)
\end{aligned}
$$

In (4), the first term is the manufacturer's expected buyback credit to the retailer, and the second term is the manufacturer's expected rebate, in which late-delivered quantity is 0 with probability α (in Case 1), $Q - Q_e$ with probability $(1 - \alpha)\beta$ (in Case 2), and Q with probability $(1 - \alpha)(1 - \beta)$ (in Case 3). Recall from

our discussion of the D-QF contract that there are two objectives in the contract design that may be conflicting with each other under the D-QF contract. Interestingly, under a BLR contract, the two objectives complement each other: On one hand, the retailer's marginal utility from the transfer payment $\partial T_{BLR}(Q, Q_e; b, \rho)/\partial Q = b(1 - \partial E[Z(Q, Q_e)]/\partial Q) + \rho w(1-\alpha)$ increases in both b and ρ. On the other hand, the manufacturer's marginal utility from the transfer payment $\partial[-T_{BLR}(Q, Q_e; b, \rho)]/\partial Q_e = b \cdot \partial E[Z(Q, Q_e)]/\partial Q_e + \rho w(1-\alpha)\beta$ also increases in both b and ρ. This suggests that whereas a high buyback price helps overcome double marginalization, it also reduces the necessity of using a high rebate level; likewise, a high rebate level helps incentivize early production, and at the same time it reduces the necessity of using a high buyback price. The proposition below details the properly designed BLR contract that coordinates the supply chain.

PROPOSITION 3. *The BLR contract coordinates the supply chain if and only if*

$$b^*_{BLR} = \frac{\beta w - c_e}{\beta(p - c_o) - c_e} \cdot p \quad and$$

$$\rho^*_{BLR} = \frac{(p - w - c_o)(c_e - \beta c_r)}{w(1-\alpha)[\beta(p - c_o) - c_e]}.$$

Under the contract, the manufacturer's and the retailer's profit shares are $[\beta(p - w - c_o)]/[\beta(p - c_o) - c_e]$ *and* $(\beta w - c_e)/[\beta(p - c_o) - c_e]$, *respectively.*

One can verify that $b^*_{BLR} < w + c_o$ holds so that the retailer will not profit from leftover inventory. To ensure that $b^*_{BLR} > 0$ and $0 < \rho^*_{BLR} \leq 1$, we need two mild conditions in addition to $c_e/\beta \leq c_r/\alpha$ (cf. Proposition 2). First, we require that $p > w + c_o$, meaning that the retail price is higher than the wholesale price plus the per-unit administration cost. Second, $\beta w > c_e$, that is, the manufacturer's expected per-unit revenue from early production (βw) is higher than the unit cost (c_e).

In §4.2, we have shown that although there exists a coordinating LR contract under a wholesale price $w = c_e/\beta$, such a contract does not allow a flexible profit division between the manufacturer and the retailer; in fact, as w deviates from c_e/β, the supply chain performance deteriorates. Under the optimal BLR contract, by Proposition 3, the retailer's profit share is $(\beta w - c_e)/[\beta(p - c_o) - c_e]$. Because the range of w is $c_e/\beta < w < p - c_o$ (see the previous paragraph), the retailer's profit share can be any value between zero and one. Therefore, the supply chain profit can be arbitrarily divided between the firms by adjusting the wholesale price w.

The following corollary provides comparative statics to show the impact of α, β, and γ on the optimal BLR contract parameters b^*_{BLR} and ρ^*_{BLR}.

COROLLARY 1. *Under a coordinating BLR contract, the following results hold*:

	b^*_{BLR}	ρ^*_{BLR}	Retailer's profit share	Manufacturer's profit share
α	—	↑	—	—
β	↑	↓	↑	↓
γ	—	—	—	—

As α increases, the regular production mode becomes more reliable. Thus, a higher rebate rate (ρ^*_{BLR}) is needed to motivate the manufacturer to operate in the early production mode, but the buyback price (b^*_{BLR}) remains unchanged because α does not affect the expected additional cost of early production ($c_e - \beta c_r$). The profit division of the supply chain between the two firms is directly tied to the buyback price b^*_{BLR} and thus remains unchanged. As β increases, early production becomes less risky. Even a lower rebate rate (ρ^*_{BLR}) can now provide the manufacturer with adequate incentive. To encourage the retailer to place a larger order, the manufacturer offers a higher buyback price (b^*_{BLR}), which gives the retailer a higher profit share. Interestingly, the optimal contract parameters are independent of the time sensitivity of demand (captured by γ). To understand why, note that γ does not affect the quantity of late-delivered products. In addition, although γ affects the quantity of the leftover inventory, it affects neither the manufacturer's nor the retailer's marginal loss or gain from the leftover inventory under the BLR contract. This remains true even when γ is a function of the on-time-delivered quantity.

6. Evaluation of Contract Performance

We now evaluate the performance of the sample contracts used by Manufacturers A and B during the period 2009–2011, as shown in Table 1. Our performance evaluation is based on the U.S. influenza vaccine market and primarily serves as an illustration of the efficiency of the influenza vaccine supply chain under different contracts. We use the actual wholesale prices for 0.5 ml syringe offered by Manufacturers A and B during the 2009–2010 season. (For confidentiality, we do not report the actual prices here. The two wholesale prices are roughly between $8 and $10, and differ only slightly.) When Manufacturer B used the wholesale price contract during the 2008–2009 and 2010–2011 seasons, its wholesale price was lower than that of Manufacturer A by 15% on average. Thus, for fair comparison between the LR contract and the wholesale price contract, we use the wholesale price for the wholesale price contract that is lower by 15% than that for the LR contract.

As is common in the supply contract literature, the demand ξ is assumed to follow a gamma distribution with a density of $f(\xi) = (\beta_g)^{-\alpha_g} e^{-\xi/\beta_g} \xi^{\alpha_g-1}/\Gamma(\alpha_g)$. In practice, we have observed that manufacturers use the same contracts for different retailers. Because we consider a retailer of any size, we normalize the mean of ξ to 1 and use 0.5 for its coefficient of variation; this leads to the shape and scale parameters: $\alpha_g = 4$ and $\beta_g = 0.25$. (We have also conducted additional numerical experiments under different parameters of the gamma distribution, and under different distributions such as uniform and normal distributions, and found that our main insights remain unchanged.) In addition, we choose $c_e = c_r = \$3$ and $p = \$18$ (see Online Appendix C).

Our numerical results summarized in Table 3 provide the following implications:

(i) The D-QF contract performs reasonably well with the average supply chain efficiency of 97.58% and the manufacturer's profit share of 42.97% under the specified contract prices. However, Manufacturer A should consider using higher return allowances to improve its performance: Our further analysis reveals

that on average $(\kappa_1^*, \kappa_2^*) = (47.7\%, 57.6\%)$ under optimal D-QF contracts, yielding a supply chain efficiency of 99.71%. On the other hand, if the Manufacturer A is solely concerned about maximizing its own profit, then we find that it is optimal to set $\kappa_1^* = 0\%$ and $\kappa_2^* = 92.78\%$ with the average wholesale price. In fact, we observed that another manufacturer in the industry used the D-QF contract with $(\kappa_1^*, \kappa_2^*) = (0, 100\%)$ during the 2009–2010 season.

(ii) Manufacturer B may consider using a BLR contract instead of using a wholesale price contract or an LR contract, under which the supply chain efficiency is, on average, 84.45% and 87.86%, respectively. A properly designed BLR contract not only guarantees channel coordination (i.e., 100% supply chain efficiency) but also provides flexibility in profit division. Compared to an LR contract, implementing a BLR contract requires returning leftover inventory, but the additional administrative burden is minimal.

7. Extensions
We now discuss three extensions to our model. Section 7.1 extends our supply chain analysis to social

Table 3   Performance of the Supply Contracts Used by Manufacturers A and B

$\{\alpha, \beta, \gamma\}$	D-QF used by A (2008–2011)			LR used by B (2009–2010)			Wholesale price contract used by B (2008–2009; 2010–2011)		
	M profit share %	R profit share %	SC efficiency %	M profit share %	R profit share %	SC efficiency %	M profit share %	R profit share %	SC efficiency %
{0.6, 0.9, 0.1}	42.49	57.51	97.46	45.52	54.48	90.09	41.08	58.92	91.35
{0.6, 0.9, 0.3}	42.94	57.06	98.25	45.60	54.40	90.53	41.48	58.52	83.30
{0.6, 0.9, 0.5}	44.03	55.97	98.75	45.84	54.16	90.47	42.82	57.18	72.50
{0.6, 0.95, 0.1}	43.50	56.50	97.79	46.34	53.66	90.20	41.08	58.92	90.14
{0.6, 0.95, 0.3}	44.04	55.96	98.23	46.37	53.63	90.30	41.48	58.52	81.44
{0.6, 0.95, 0.5}	45.18	54.82	98.69	46.50	53.50	90.23	42.82	57.18	70.38
{0.6, 0.99, 0.1}	44.29	55.71	97.77	46.97	53.03	90.02	40.89	59.11	92.56
{0.6, 0.99, 0.3}	44.86	55.14	98.07	46.98	53.02	90.02	40.89	59.11	92.56
{0.6, 0.99, 0.5}	46.07	53.93	98.60	47.00	53.00	90.00	40.92	59.08	92.54
{0.7, 0.9, 0.1}	42.37	57.63	97.00	45.43	54.57	90.24	41.07	58.93	92.02
{0.7, 0.9, 0.3}	42.69	57.31	97.87	45.79	54.21	84.34	41.45	58.55	86.10
{0.7, 0.9, 0.5}	43.49	56.51	98.37	47.04	52.96	76.03	42.72	57.28	77.93
{0.7, 0.95, 0.1}	43.33	56.67	97.48	46.36	53.64	90.13	41.07	58.93	91.00
{0.7, 0.95, 0.3}	43.72	56.28	97.90	46.39	53.61	90.28	41.45	58.55	84.38
{0.7, 0.95, 0.5}	44.52	55.48	98.28	46.49	53.51	90.25	42.72	57.28	75.92
{0.7, 0.99, 0.1}	44.09	55.92	97.52	46.98	53.02	90.02	41.07	58.93	89.88
{0.7, 0.99, 0.3}	44.50	55.50	97.76	46.98	53.02	90.02	41.45	58.55	82.93
{0.7, 0.99, 0.5}	45.34	54.66	98.17	47.00	53.00	90.00	42.72	57.28	74.32
{0.8, 0.9, 0.1}	42.64	57.36	97.69	46.00	54.00	90.41	41.07	58.93	92.45
{0.8, 0.9, 0.3}	42.46	57.54	97.42	46.29	53.71	86.90	41.36	58.64	88.96
{0.8, 0.9, 0.5}	42.95	57.05	97.87	47.30	52.70	81.42	42.37	57.63	83.61
{0.8, 0.95, 0.1}	43.18	56.82	97.12	46.39	53.61	90.01	41.07	58.93	91.82
{0.8, 0.95, 0.3}	43.42	56.58	97.54	46.41	53.59	90.24	41.36	58.64	87.45
{0.8, 0.95, 0.5}	43.92	56.08	97.84	46.48	53.53	90.26	42.37	57.63	81.78
{0.8, 0.99, 0.1}	43.89	56.11	97.27	46.98	53.02	90.01	41.07	58.93	90.83
{0.8, 0.99, 0.3}	44.16	55.84	97.43	46.99	53.01	90.02	41.36	58.64	86.10
{0.8, 0.99, 0.5}	44.68	55.32	97.71	47.00	53.00	90.01	42.37	57.63	80.28
Average	43.80	56.20	97.85	46.50	53.50	88.98	41.61	58.39	85.35

Note. "M" stands for the manufacturer, "R" for the retailer, and "SC" for the supply chain.

welfare. Section 7.2 considers the case where the manufacturer produces only one monovalent vaccine early with a strain that is most likely to be selected by the Committee. Section 7.3 discusses the extension with random production yield.

7.1. Social Welfare

Motivated by the real contracts used in practice between vaccine manufacturers and healthcare providers (retailers), we have so far focused our attention on how a contract between these two parties coordinates the firms' incentives to maximize their total profits. Since the influenza vaccine affects public health, we now bring consumers' benefits into context and analyze the impact of supply contract design on social welfare that is commonly defined as the sum of consumer surplus and the supply chain's profit. Since social welfare includes the supply chain's profit, the analysis in this section will build on our earlier supply chain analysis.

We model consumers' benefits by estimating the infection probability of the population. Without loss of generality, we normalize the size of the population to one. Then, the ex-post sales quantity represents the proportion of vaccinated population, denoted by z; i.e., $z = Z(Q, Q_e \mid \xi)$. We let z_1 denote the number of doses administered during the ideal period, $z_2 = z - z_1$ the number of doses administered after the ideal period, and $\phi \in (0, 1)$ the efficacy of vaccine. Then, using an SIR Model with vital dynamics (Adida et al. 2013, Keeling and Rohani 2008, Mamani et al. 2012), we show in Online Appendix B that an individual's overall probability of infection during the influenza vaccine season is $1 - \phi(z - \tau z_2) - 1/R_0$, where τ denotes the ratio of the length of the ideal vaccination period over that of the influenza season, and R_0 is the basic reproductive ratio, a constant that measures the expected number of individuals infected by each infectious individual. Letting l denote the average financial loss incurred by an individual from infection (similar to Chick et al. 2008, Yamin and Gavious 2013), the consumers' benefits from vaccination can be captured by the resultant reduction in financial loss, that is, $-l[1 - \phi(z - \tau z_2) - 1/R_0] - [-l(1 - 1/R_0)] = l\phi(z - \tau z_2)$. Finally, the expected social welfare, denoted by $W(Q, Q_e)$, is represented as the sum of expected consumer surplus and expected supply chain profit:
$$W(Q, Q_e) = l\phi\{E[Z(Q, Q_e)] - \tau E[Z_2(Q, Q_e)]\} - [(c_r + c_o)Q + (c_e - \beta c_r)Q_e].$$

In this setting, the following corollary shows that the government's subsidy to a *retailer*, together with a well-calibrated supply contract, can align the supply chain members' incentives to achieve the social optimum.

Corollary 2. *A hybrid incentive scheme, in which on top of a coordinating supply contract between the manufacturer and the retailer, a third party such as the government provides the retailer with a unit subsidy of $(l\phi - p)$ for the sales generated during the ideal vaccination period and a unit subsidy of $[l\phi(1 - \tau) - p]$ for the sales generated after the ideal vaccination period, can induce the maximum social welfare.*

Under the proposed subsidy scheme, the retailer's total revenue matches the social welfare impact from the vaccination, and therefore the subsidy to the retailer together with a coordinating supply contract between the manufacturer and the retailer can achieve the social optimum. The notion of introducing third-party intervention is not new. It has been proposed in the literature (e.g., Adida et al. 2013, Arifoğlu et al. 2012, Chick et al. 2008) that the government might subsidize or share costs based on the total number of vaccines produced by a *manufacturer* to bring the vaccine market closer to the social optimum. However, such results are obtained in the setting where the demand is deterministic and the government determines the demand for vaccine (Chick et al. 2008) or the manufacturer sells vaccines directly to consumers (e.g., Adida et al. 2013, Arifoğlu et al. 2012). In the U.S. market, however, retailers place orders to manufacturers and then distribute vaccines to consumers. One implication from Corollary 2 is that contrary to the existing results in the literature, it is not possible to achieve the social optimum by providing a per-unit subsidy only to the manufacturer based on the early production quantity Q_e (or the total production quantity Q). To understand why, suppose that the government provides a per-unit subsidy to the manufacturer based on Q_e. Then the manufacturer may have a lower early production cost and increase Q_e. However, the subsidy scheme would not increase the retailer's order quantity Q to the socially optimal level unless the government provides another subsidy to the retailer as well. On the other hand, our result indicates that to achieve the social optimum it is sufficient to use a coordinating contract together with a subsidy to the retailer. This novel insight is obtained in our model by taking into account the important role of retailers in the U.S. influenza vaccine supply chain that have been neglected in prior literature.

7.2. Trivalent Vaccine Under Continuous Distribution of On-Time Delivery

As in the literature reviewed in §2, our base model assumes away the consideration that a typical influenza vaccine consists of three virus strains, one from each of subtypes A/H1N1, A/H3N2, and type B. In practice, a vaccine manufacturer first produces each monovalent vaccine that contains one strain, and then combines three monovalent vaccines into a trivalent vaccine. As such, a manufacturer usually initiates at-risk early production of one monovalent vaccine with a strain that is most likely (but still uncertain) to be selected by the Committee (VRBPAC 2002–2014). Moreover, our

base model captures the uncertainty in delivery by assuming that all vaccines from regular production will be delivered either on time (with probability α) or late (with probability $1 - \alpha$). We relax this assumption by using a continuous random variable, denoted by $\chi \in [0, 1]$, that represents a random proportion of on-time delivered units from regular production. We denote by $\bar{\chi}$ the mean of χ.

Under this new setting, we maintain the same sequence of events as in the base model but add the following reinterpretations: (1) $t = 1$: Q_e represents an early-production quantity of monovalent vaccine containing a strain that is most likely to be selected by the Committee but is still uncertain. (2) $t = 2$: If early production is successful with probability β, then the manufacturer produces the remaining $Q_r = Q - Q_e$ units of the monovalent vaccine and then produces Q units of each of two other monovalent vaccines. Otherwise, the manufacturer produces Q units of each of three monovalent vaccines containing the three strains chosen by the Committee. (3) $t = 3$ to $t = 4$: The manufacturer delivers $JQ_e + \chi Q_r$ units of trivalent vaccines (final products) on time during the ideal period ($t = 3$) and delivers $(1 - \chi)Q_r$ units of trivalent vaccines in the late period ($t = 4$).

COROLLARY 3. *In the setting described above, a BLR contract with $b^*_{BLR} = ((\beta w - c_e)/(\beta(p - c_o) - c_e)) \cdot p$ and $\rho^*_{BLR} = (p - w - c_o)(c_e - \beta c_r)/(w(1 - \bar{\chi})[\beta(p - c_o) - c_e])$ coordinates the supply chain.*

We can also show that there exist exogenous parameters such that no D-QF contract coordinates the supply chain and that a coordinating LR contract exists only when $w = c_e/\beta$ and the retailer takes all the profit of the supply chain.

7.3. Random Yield

The uncertainty associated with the production of influenza vaccine has two dimensions: delivery timing and output quantity caused by random yield. The first dimension has been one of the primary causes of mismatch between demand and supply but has not yet been studied in the literature. Thus, our paper focuses on the first dimension while suppressing the second dimension to maintain tractability. Without considering random yield, our base model presented in §3 assumes that the manufacturer's total production quantity from both early and regular production modes equals the retailer's order quantity Q. Now consider the case with yield uncertainty. Let θ denote a random yield with a support of $[\underline{\theta}, \bar{\theta}]$, a density $h(\cdot)$, and a cumulative distribution $H(\cdot)$. When the production quantity targeted by the manufacturer (or the number of chicken eggs put into production) is n, the quantity of vaccines produced is θn. In this case, the manufacturer may choose its target production quantity differently

from the retailer's order quantity Q. Unfortunately, with four different sources of uncertainties (design, delivery, demand, and yield) in one model, it is very challenging to characterize the optimal target production quantity. Thus, for tractability, we assume that the manufacturer chooses a target production quantity in proportion to the retailer's order quantity; i.e., Q/k for some fixed constant $k > 0$. This assumption is consistent with the literature; for example, Chick et al. (2008) derive k endogenously from the yield distribution (Corollary 1), and Federgruen and Yang (2009) choose $k = E[\theta]$. In addition, Choi et al. (2008), Güler and Bilgiç (2009), and Gurnani et al. (2000), among others, assume that the buyer determines the supplier's target (input) production quantity.

The sequence of events under random yield is similar to that of the base model except for the following: At $t = 1$, after receiving the retailer's order Q, the manufacturer determines the target quantity n_e for early production. At $t = 2$, upon the release of the finalized product design, the manufacturer determines the target quantity n_r for regular production according to $n_r = Q/k - Jn_e$. At $t = 3$, the manufacturer observes the realized yield θ. The quantity of vaccines from early production—if successful—shipped to the ideal period is θn_e, and the quantity of vaccines from regular production is θn_r. In this setting, we can replicate the following results from the base model without random yield (see Online Appendix A for abridged proofs): (1) None of the conventional contracts, including wholesale price, buyback, and QF contracts, can coordinate the supply chain. As the variance of θ increases, these contracts become less efficient. (2) The D-QF contract may coordinate the supply chain only under a set of conditions, and its parameters are affected by both the mean and variance of θ. (3) The LR contract coordinates the supply chain only under a specified wholesale price, and the coordinating contract parameters are not affected by the variance of θ. (4) A properly designed BLR contract can always coordinate the supply chain, and its parameters are affected by the mean but not the variance of θ.

To demonstrate the robustness of our results, we have conducted numerical experiments in which the manufacturer determines k endogenously. When using the same set of parameters as in Table 3 under random yield following a uniform distribution between 0 and 2, we found that supply chain efficiency drops slightly but the relative performance among the contracts remains the same.

8. Concluding Remarks

In this paper we study a contract design problem for the U.S. influenza vaccine supply chain that faces uncertainties in the design, delivery, and demand of

the product. To mitigate the risk of late delivery, the manufacturer operates in two production modes: the regular production mode that starts production after the design uncertainty is resolved and the early production mode that starts production before the design is finalized. The manufacturer faces a trade-off between the informational advantage of regular production and the delivery advantage of early production. Currently, the industry is experimenting with different contract types, aiming to improve its supply chain performance. This makes our study both explanatory and prescriptive.

Our analysis reveals that without proper contract design, a vicious incentive cycle may arise: Because the manufacturer bears the risk associated with early production, it lacks the incentive to improve on-time delivery, which reduces the retailer's order size in anticipation of lost sales; this further discourages the manufacturer from making an effort to improve its delivery performance. As a result, conventional contracts fail to coordinate this supply chain. We proceed to analyze the two contract forms that are commonly used in practice: the delivery-time-dependent quantity flexibility (D-QF) contract and the late-rebate (LR) contract. We find that both contracts may fail to coordinate the supply chain under realistic settings because of the tension between overcoming double marginalization and incentivizing early production. These findings indicate that a careful analysis of the interaction of multiple uncertainties is needed in contract design—a haphazard addition of more complexity to a contract does not necessarily lead to coordination. By leveraging the strengths while overcoming the limitations of the D-QF and LR contracts, we construct a buyback-and-late-rebate (BLR) contract, which can not only coordinate the supply chain but also can provide flexibility of profit division. These insights can help practitioners design supply contracts for improving on-time delivery performance of the influenza vaccine supply chain and potentially do the same for other supply chains with similar characteristics.

As for future research, studying sophisticated contracts and negotiations between multiple manufacturers and multiple distributors/retailers in a dynamic setting will be an interesting research avenue. This has been only partially addressed in the literature; e.g., Federgruen and Yang (2008) study the problem of a retailer selecting a portfolio of manufacturers under a fixed wholesale price, and Cho and Tang (2013) study dynamic ordering and selling under a wholesale price contract between a manufacturer and a retailer. Our analysis of various contracts presented in this paper will be useful building blocks to further advance this line of research.

Supplemental Material

Acknowledgments

References

Adida E, Dey D, Mamani H (2013) Operational issues and network effects in vaccine markets. *Eur. J. Oper. Res.* 231(2):414–427.

Arifoğlu K, Deo S, Iravani S (2012) Consumption externality and yield uncertainty in the influenza vaccine supply chain: Interventions in demand and supply sides. *Management Sci.* 58(6):1072–1091.

Arya A, Mittendorf B (2004) Using return policies to elicit retailer information. *RAND J. Econom.* 35(3):617–630.

Cachon G (2003) Supply chain coordination with contracts. De Kok AG, Graves SC, eds. *Handbooks in Operations Research and Management Science*, Supply Chain Management: Design, Coordination, and Operation, Vol. 11 (Elsevier Science, Amsterdam), 229–340.

Cachon G, Lariviere M (2001) Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Sci.* 47(5):629–646.

Cachon GP, Zhang F (2006) Procuring fast delivery: Sole-sourcing with information asymmetry. *Management Sci.* 52(6):881–896.

Centers for Disease Control and Prevention (CDC) (2014) Seasonal influenza vaccine supply and distribution in the United States. Accessed March 29, 2015, http://www.cdc.gov/flu/about/qa/ vaxdistribution.htm.

Chick SE, Hasija S, Nasiry J (2012) Incentive alignment and information elicitation from manufacturers for public goods procurement. Working paper, INSEAD, Fontainebleau, France.

Chick SE, Mamani H, Simchi-Levi D (2008) Supply chain coordination and influenza vaccination. *Oper. Res.* 56(6):1493–1506.

Cho S-H (2010) The optimal composition of influenza vaccines subject to random production yields. *Manufacturing Service Oper. Management* 12(2):256–277.

Cho S-H, Tang CS (2013) Advance selling in a supply chain under uncertain supply and demand. *Manufacturing Service Oper. Management* 15(2):305–319.

Choi HCP, Blocher JD, Gavirneni S (2008) Value of sharing production yield information in a serial supply chain. *Production Oper. Management* 17(6):614–625.

Deo S, Corbett CJ (2009) Cournot competition under yield uncertainty: The case of the U.S. influenza vaccine market. *Manufacturing Service Oper. Management* 11(4):563–576.

Deo S, Kolesov I, Waikar S (2012) Optimizing flu vaccine planning at NorthShore University HealthSystem. Teaching case, Kellogg School of Management, Northwestern University, Evanston, IL.

Donohue KL (2000) Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Sci.* 46(11):1397–1411.

Federgruen A, Yang N (2008) Selecting a portfolio of suppliers under demand and supply risks. *Oper. Res.* 56(4):916–936.

Federgruen A, Yang N (2009) Competition under generalized attraction models: Applications to quality competition under yield uncertainty. *Management Sci.* 55(12):2028–2043.

Fukuda K, O'Mara D, Singleton JA (2002) How the delayed distribution of influenza vaccine created shortages in 2000 and 2001. *Pharmacy Therapeutics* 27(5):235–242.

Grout JR, Christy DP (1993) An inventory model of incentives for on-time delivery in just-in-time purchasing contracts. *Naval Res. Logist.* 40(6):863–877.

Güler MG, Bilgiç T (2009) On coordinating an assembly system under random yield and random demand. *Eur. J. Oper. Res.* 196(1):342–350.

Gurnani H, Akella R, Lehoczky J (2000) Supply management in assembly systems with random yield and random demand. *IIE Trans.* 32(8):701–714.

Hwang W, Bakshi N, DeMiguel V (2014) Simple contracts for reliable supply: Capacity versus yield uncertainty. Working paper, London Business School, London.

Keeling MJ, Rohani P (2008) *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ).

Lariviere M (1998) Supply chain contracting and coordination with stochastic demand. Tayur S, Ganeshan R, Magazine M, eds. *Quantitative Models for Supply Chain Management* (Kluwer, Boston), 235–268.

Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN (2007) Influenza seasonality: Underlying causes and modeling theories. *J. Virology* 81(11):5429–5436.

Loftus P (2014) Flu-shot supply dented by production woes. *Wall Street Journal* (October 26) B7.

Mamani H, Adida E, Dey D (2012) Vaccine market coordination using subsidy. *IIE Trans. Healthcare Systems Engrg.* 2(1):78–96.

Matthews JT (2006) Egg-based production of influenza vaccine: 30 years of commercial experience. *Bridge* 36(3):17–24.

Nowalk MP, Zimmerman RK, Cleary SM, Bruehlman RD (2005) Missed opportunities to vaccinate older adults in primary care. *J. Amer. Board Family Medicine* 18(1):20–27.

O'Mara D, Fukuda K, Singleton JA (2003) Influenza vaccine: Ensuring timely and adequate supply. *Infections Medicine J.* 20(11):548–554.

Padmanabhan V, Png IPL (1997) Manufacturer's returns policies and retail competition. *Marketing Sci.* 16(1):81–94.

Pasternack BA (1985) Optimal pricing and return policies for perishable commodities. *Marketing Sci.* 4(2):166–176.

Taylor TA (2002) Supply chain coordination under channel rebates with sales effort effects. *Management Sci.* 48(8):992–1007.

Tsay A (1999) Quantity-flexibility contract and supplier-customer incentives. *Management Sci.* 45(10):1339–1358.

Vaccines and Related Biological Products Advisory Committee (VRBPAC) (2002–2014) Additional information. Accessed November 17, 2015, http://www.fda.gov/cber/advisory/vrbp/vrbpmain.htm.

Yamin D, Gavious A (2013) Incentives' effect in influenza vaccination policy. *Management Sci.* 59(12):2667–2686.