

# Maximizing the effectiveness of a pediatric vaccine formulary while prohibiting extrimmunization

Shane N. Hall · Edward C. Sewell · Sheldon H. Jacobson

Received: 25 May 2007 / Accepted: 16 October 2007 / Published online: 7 May 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** The growing complexity of the United States Recommended Childhood Immunization Schedule has resulted in as many as five required injections during a single well-baby office visit. To reduce this number, vaccine manufacturers have developed combination vaccines that immunize against several diseases in a single injection. At the same time, a growing number of parents are challenging the safety and effectiveness of vaccinating children. They are also particularly concerned about the use of combination vaccines, since they believe that injecting a child with multiple antigens simultaneously may overwhelm a child's immune system. Moreover, combination vaccines make it more likely that extrimmunization (i.e., administering more than the required amount of vaccine antigens) occurs, resulting in greater concerns by parents and vaccine wastage costs borne by an already strained healthcare system. This paper formulates an integer programming model that solves for the maximum number

of vaccines that can be administered without any extrimmunization. An exact dynamic programming algorithm and a randomized heuristic for the integer programming model is formulated and the heuristic is shown to be a randomized  $\xi$ -approximation algorithm. Computational results are reported on three sets of test problems, based on existing and future childhood immunization schedules, to demonstrate their computational effectiveness and limitations. Given that future childhood immunization schedules may need to be solved for each child, on a case-by-case basis, the results reported here may provide a practical and valuable tool for the public health community.

**Keywords** Pediatric immunization · Pediatric vaccines · Randomized approximation algorithms · Discrete optimization · Dynamic programming · Heuristics · Integer programming

---

S. N. Hall (✉)  
Department of Operational Sciences,  
Air Force Institute of Technology,  
2950 Hobson Way, Wright-Patterson AFB,  
OH 45433-7765, USA  
e-mail: shane.hall@afit.edu

E. C. Sewell  
Department of Mathematics and Statistics,  
Southern Illinois University,  
Edwardsville, IL 62026-1653, USA  
e-mail: esewell@siue.edu

S. H. Jacobson  
Simulation and Optimization Laboratory,  
Department of Computer Science, University of Illinois,  
201 North Goodwin Avenue(MC- 258),  
Urbana, IL 61801-2302, USA  
e-mail: shj@uiuc.edu

## 1 Introduction

Parental refusal to vaccinate their child has become a growing concern for public health administrators and pediatricians. For example, in a recent national survey of pediatricians, 54% had encountered parents over a 12-month period that refused to vaccinate their child, citing safety concerns as their top reason [14]. In another survey, 70% of pediatricians had encountered a parent in the 12 months preceding the survey that refused at least one immunization for their child [12]. As these studies indicate, the safety, necessity, and effectiveness of vaccines are a concern for many parents.

In the last decade, vaccine manufacturers have begun developing a growing number of combination vaccines, which protect against several diseases with a single

injection [13, 18]. Such combination vaccines are an important solution to alleviating vaccine injection overcrowding and schedule complexity [8, 13, 31] within the current United States Recommended Childhood Immunization Schedule [4]. However, using combination vaccines increases the risk of extrimmunization, which means that a child receives vaccine antigens for a given disease beyond what is prescribed in the United States Recommended Childhood Immunization Schedule. For example, injecting a child with a diphtheria, tetanus, pertussis, hepatitis B, and polio combination vaccine at age 4 months would provide extrimmunization for hepatitis B, since no dose of this vaccine is required at that age (see Fig. 1). Such extrimmunization poses biological safety risks and amplifies parental concerns on vaccine safety and effectiveness. Biologically, extrimmunization of some vaccines increases the risk of adverse side effects; such is the case with diphtheria and tetanus vaccines [8]. Furthermore, several people challenge the safety and effectiveness of vaccinating children and particularly object to the use of combination vaccines, since they believe that injecting a child with multiple antigens simultaneously overwhelms the infant immune system, and hence, extrimmunization due to combination vaccines only exacerbates these fears [10, 13]. Moreover, the economic toll of extrimmunization is also significant. For example, the annual societal cost burden of providing one extra dose of vaccine for each child born in the United States is no less than \$28 million, which assumes a birth rate of 11,100 births per day (see [26]) and a vaccine cost of \$7, where the vaccine cost estimates the Federal contract purchase price of the least expensive pediatric vaccine (see [3]).

In addition to extrimmunization, combination vaccines are creating a combinatorial explosion of vaccine alternatives and choices for the public health community, especially as the Recommended Childhood Immunization Schedule continues to evolve, with new diseases emerging

and/or new vaccines being developed. For example, four time periods and five diseases (with Rotavirus and Meningococcal added in 2007) have been added to the Recommended Childhood Immunization Schedule since 1995 [5, 9, 11]. This combinatorial explosion of vaccine alternatives and choices has motivated the use of operations research techniques in addressing pediatric immunization formulary design problems. Most of the research to date addresses the economics of pediatric vaccine formulary design, combination vaccine pricing, and vaccine wastage [21, 23]. Weniger et al. [32] report the results of a pilot study that uses operations research methods to assess the economic value of vaccine formularies (i.e., the set of vaccines inventoried by an immunization clinic or pediatrician). In this pilot study, a portion of the Recommended Childhood Immunization Schedule was modeled as an integer program (IP). The objective of this IP was to aide decision-makers in determining the vaccine formulary that minimized the cost to fully immunize a child. Jacobson et al. [26] present a more rigorous presentation of this pilot study and demonstrate how the model selects different vaccine formularies depending on the desired economic criteria. Sewell et al. [30] embed the IP from the pilot study into a bisection algorithm [2] to “reverse engineer” the maximum inclusion prices (the maximum price at which a vaccine remains part of the optimal vaccine formulary) of four combination vaccines not yet (at that time) licensed in the United States. Sewell and Jacobson [29] present a rigorous description of this study, including the complete IP model. This study shows how operations research can provide beneficial economic analysis to the pharmaceutical companies that develop and manufacture vaccines (see [20, 25] for additional applications of this bisection algorithm). Jacobson and Sewell [22] extend the bisection/IP algorithm approach by including it with Monte Carlo simulation, thereby determining a probability distribution for the price

DISEASE	TIME PERIOD (Age of Child)									
	1 (Birth)	2 (1 Mo)	3 (2 Mos)	4 (4 Mos)	5 (6 Mos)	6 (12 Mos)	7 (15 Mos)	8 (18 Mos)	9 (24 Mos)	10 (4-6 Yrs)
Hepatitis B	D1	D2			D3					
Diphtheria, Tetanus, Pertussis			D1	D2	D3		D4			D5
<i>Haemophilus influenzae</i> type b			D1	D2	D3	D4				
Polio			D1	D2	D3					D4
Measles, Mumps, Rubella						D1				D2
Varicella						D1				
Pneumococcus			D1	D2	D3	D4				
Influenza					D1 (yearly)					
Hepatitis A						D1		D2		

Fig. 1 United States 2006 recommended childhood immunization schedule through age 6 (excluding recommendations for selected populations)

of the four potential combination vaccines. Lastly, Hall et al. (submitted) formulate and analyze a general discrete optimization model that determines the set of vaccines (i.e., a vaccine formulary) that should be used in a clinical environment to satisfy any given childhood immunization schedule at minimum cost, while Hall [16] formulates and analyzes a general discrete optimization model that determines the set of vaccines that should be used in a clinical environment to satisfy any given childhood immunization schedule that minimizes the amount of extraimmunization.

This paper addresses the safety of combination vaccines by further examining the issue of extraimmunization as an extension of the results in Hall [16]. Specifically, this paper presents a general model that maximizes the effectiveness of a given pediatric vaccine formulary while also prohibiting extraimmunization for any given childhood immunization schedule. The paper is organized as follows. Section 2 presents the model terminology and the formulation of the discrete optimization problem. Section 3 presents a description and analysis of an exact dynamic programming algorithm and a randomized heuristic for solving the discrete optimization problem. Section 4 presents a computational comparison of this dynamic programming algorithm and randomized heuristic with an integer programming branch and bound algorithm, while Section 5 provides concluding comments and directions for future research.

## 2 Model terminology and formulation

The model terminology and the formulation of a discrete optimization problem used to maximize the effectiveness of a given pediatric vaccine formulary while prohibiting extraimmunization for any given childhood immunization schedule is presented. Some extensions of the discrete optimization problem are also described.

The discrete optimization model assumes a given childhood immunization schedule and vaccine formulary. To precisely describe a childhood immunization schedule, several set and parameter definitions are needed. Let

- $T = \{1, 2, \dots, \tau\}$  be the set of time periods for the childhood immunization schedule,
- $D = \{1, 2, \dots, \delta\}$  be the set of diseases requiring immunization,
- $D_E \subseteq D$ , with  $|D_E| = \delta_E$  be the set of diseases where extraimmunization is permitted,
- $D_{NE} = D \setminus D_E$ , with  $|D_{NE}| = \delta_{NE}$  be the set of diseases where extraimmunization is not permitted,
- $V = \{1, 2, \dots, v\}$  be the vaccine formulary (i.e., set of vaccines available to be administered to immunize against the  $\delta$  diseases),

- $n_d \in \mathbb{Z}^+$  be the number of doses of a vaccine that must be administered for immunization against disease  $d \in D$
- $I_{vd} = 1$  if vaccine  $v \in V$  immunizes against disease  $d \in D$ , 0 otherwise, which is a set of binary parameters that indicate which vaccines immunize against which diseases,
- $P_{djt} = 1$  if in time period  $t \in T$ , a vaccine may be administered to satisfy the  $j$ th dose,  $j = 1, 2, \dots, n_d$ , requirement for disease  $d \in D$ , 0 otherwise, which is a set of binary parameters that indicate the set of time periods in which a particular dose of a vaccine may be administered to immunize against a disease, and, finally,
- $R_{dt} = 1$  if in time period  $t \in T$ , no dose of a vaccine may be administered to immunize against disease  $d \in D_{NE}$ , 0 otherwise, which is a set of binary parameters that indicate the set of time periods in which no dose of a vaccine may be administered to immunize against a disease where extraimmunization is not permitted.

These sets and parameters completely describe a childhood immunization schedule along with the vaccine formulary. For example, the United States Recommended Childhood Immunization Schedule (see Fig. 1) outlines the vaccinations required to protect a child against several (currently thirteen) infectious diseases that pose a risk to children living in the United States [4]. This schedule includes the number of required doses of each vaccine and the recommended age for each dose (D1=Dose 1, D2=Dose 2, etc.). For example, the disease  $d$ =hepatitis B requires three doses of vaccine (i.e.,  $n_d=3$ ), where the second dose (D2) may be administered at age 1 or 2 months. Therefore, for disease  $d$ =hepatitis B, dose  $j=2$ ,  $P_{djt}=1(0)$  for time periods  $t=2,3(1,4,5,6,7,8,9,10)$ , and, assuming hepatitis B  $\in D_{NE}$ ,  $R_{dt}=1(0)$  for time periods  $t=4,9,10(1,2,3,5,6,7,8)$ .

Unless otherwise stated, the phrase “childhood immunization schedule” refers to an arbitrary general immunization schedule, whereas the phrase “Recommended Childhood Immunization Schedule” refers to the immunization schedule in Fig. 1. Assume that all diseases  $d \in D$  have sequentially ordered and mutually exclusive doses (i.e., all the time periods during which dose  $j$  can be administered occur prior to all the time periods during which dose  $j+1$  can be administered,  $j=1, 2, \dots, n_d-1$ ). Furthermore, assume that the time periods during which dose  $j$  can be administered are consecutive. Define the *valency*, denoted by  $Val(v)$ , as the number of diseases covered by vaccine  $v \in V$ , and hence,  $Val(v) = \sum_{d \in D} I_{vd}$ . Combination vaccines are often referred to as *multivalent* vaccines, or simply *multivalents*, because  $Val(v) \geq 2$  when  $v \in V$  is a combination vaccine. Furthermore, vaccine  $v \in V$ , where  $Val(v) = 1, 2, 3, 4, 5$ , or 6 is often referred to as a monovalent, bivalent,

trivalent, tetravalent, pentavalent, or hexavalent vaccine, respectively. In practice, the dose parameters,  $n_d$ , depend on biological constraints and are determined by the recommendations and guidelines set forth by the Centers for Disease Control and Prevention [7]. Note that schedule parameters  $P_{djt}$  specify the time periods when vaccination is “permitted” (or useful) for disease  $d \in D$ , while the schedule parameters  $R_{dt}$  specify the time periods when vaccination is “restricted” for disease  $d \in D_{NE}$ . For example, assuming disease  $d$ =hepatitis B  $\in D_{NE}$ , Fig. 1 implies  $P_{djt}=1(0)$  for some dose  $j$  and time periods  $t=1,2,3,5,6,7,8(4,9,10)$  and  $R_{dt}=1(0)$  for time periods  $t=4,9,10(1,2,3,5,6,7,8)$ . The set  $D_{NE}$  is the set of diseases where extrimmunization is restricted based on biological and/or philosophical constraints, and hence, may change for each child, on a case-by-case basis. At present, diphtheria, tetanus, and pertussis are the only diseases for which extrimmunization is biologically prohibited [8]; however, for any parent/guardian questioning the safety, necessity, and effectiveness of all vaccines, it is reasonable to assume that extrimmunization is prohibited for all diseases  $d \in D$ .

To present the discrete optimization model formulation, define the binary decision variables:

- $X_{tv}=1$  if vaccine  $v \in V$  is administered in time period  $t \in T$ , 0 otherwise,
- $\pi_{dj}=1$  if the  $j$ th dose requirement for disease  $d \in D$  is satisfied, 0 otherwise.

Therefore, for a given childhood immunization schedule, the discrete optimization model that maximizes the effectiveness of a given pediatric vaccine formulary while prohibiting extrimmunization is the following integer program

IP

$$\text{Maximize } \sum_{d \in D} \sum_{j=1}^{n_d} \pi_{dj} \tag{O}$$

Subject to

$$\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} \geq \pi_{dj} \quad \text{for all } d \in D_E, j = 1, 2, \dots, n_d, \tag{1}$$

$$\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} = \pi_{dj} \quad \text{for all } d \in D_{NE}, j = 1, 2, \dots, n_d, \tag{2}$$

$$\sum_{t \in T} \sum_{v \in V} R_{dt} X_{tv} I_{vd} = 0 \quad \text{for all } d \in D_{NE}, \tag{3}$$

$$X_{tv} \in \{0, 1\} \quad \text{for all } t \in T, v \in V, \tag{4}$$

$$\pi_{dj} \in \{0, 1\} \quad \text{for all } d \in D, j = 1, 2, \dots, n_d. \tag{5}$$

Therefore, the  $j$ th dose requirement for disease  $d \in D$  is satisfied if and only if  $\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} \geq 1$  for  $d \in D_E$ ,  $j=1,2,\dots,n_d$  and  $\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} = 1$  for  $d \in D_{NE}$ ,  $j=1,2,\dots,n_d$ . Note that the objective function (O) of IP is maximized when  $\pi_{dj}=1$  for all diseases  $d \in D$  and doses  $j=1,2,\dots,n_d$ , and hence, if the value of the optimal solution equals  $\sum_{d=1}^{\delta} n_d$ , then the  $j$ th dose requirement for every disease  $d \in D$  is satisfied and no extra vaccine doses were administered for any disease  $d \in D_{NE}$ . Therefore, IP maximizes the number of doses that may be administered while prohibiting extrimmunization. Constraint (1) ensures that for each disease  $d \in D_E$ , if dose  $j=1,2,\dots,n_d$  is satisfied, then at least one vaccine that provides immunization for disease  $d \in D_E$  is administered in some time period when dose  $j=1,2,\dots,n_d$  may be administered. Constraint (2) ensures that for each disease  $d \in D_{NE}$ , if dose  $j=1,2,\dots,n_d$  is satisfied, then exactly one vaccine that provides immunization for disease  $d \in D_{NE}$  is administered in some time period when dose  $j=1,2,\dots,n_d$  may be administered. Constraint (3) is for each disease  $d \in D_{NE}$  and ensures that the number of doses administered in time periods when vaccination is restricted equals zero. Lastly, Constraints (4) and (5) are the binary constraints on the respective decision variables. Note that there are additional special restrictions unique to certain vaccines and manufacturers that are not modeled by IP; Hall [16] describes some of these restrictions (see [5]) and how they may be modeled. Furthermore, note that in IP if, for a given disease  $d \in D$ , dose  $j$  is satisfied without satisfying vaccine dose  $j-1$ , then dose  $j$  effectively becomes dose  $j-1$ . This is reasonable since, in practice, if a child misses vaccine doses early in the Recommended Childhood Immunization Schedule then the child continues to receive vaccine doses as recommended. For example, if doses 2 and 3 for Hepatitis B are satisfied in IP (i.e.,  $\pi_{\text{HepatitisB},2} = \pi_{\text{HepatitisB},3} = 1$ ), but not dose 1, then Hepatitis B would effectively have two of the three dose series satisfied.

Hall [16] formulates and analyzes a discrete optimization model that determines the set of vaccines (i.e., a vaccine formulary) that should be used in a clinical environment to satisfy any given childhood immunization schedule that minimizes the amount of extrimmunization. Therefore, the discrete optimization model in Hall [16] emphasizes satisfying the childhood immunization schedule first while trying to minimize extrimmunization whereas IP emphasizes prohibiting extrimmunization first while trying to satisfy the childhood immunization schedule as much as possible. Furthermore, the discrete optimization model in Hall [16] is not a binary integer program (BIP); however, all the decision variables in IP are binary, which offers theoretical advantages (see [27]). For example, IP shares a similar structure to the Max-Satisfiability Problem [17] or, when  $\delta_{NE}=\delta$ , to the Set-Partitioning Problem [27], both of which have been well-

studied. In fact, IP can be shown to be *NP*-hard using Satisfiability or Set-Partitioning [15]. See Hall [16] for a complete analysis of the computational complexity and for special cases when IP is solvable in polynomial time.

Assigning weights to the objective function decision variables in IP offers a useful practical extension. Each vaccine has a known *immunogenicity* (the ability of the vaccine to immunize against a disease). For example, the immunogenicity of a Hepatitis B vaccine is >95%, which means that at least 95% of those children receiving the three dose series develop a protective antibody response against the disease [6]. However, for some vaccines, the immunogenicity increases with each dose. For example, for a Hepatitis B vaccine, the immunogenicity is 30–50% after the first dose, 75% after the second, and >95% after the third dose [6]. Therefore, if each objective function decision variable  $\pi_{dj}$  in IP is weighted by immunogenicity, then IP maximizes the total immunogenicity of the vaccine formulary while prohibiting extraimmunization. More specifically, define  $\text{Imm}_{dj}$  as the immunogenicity for disease  $d \in D$  after dose  $j=1,2,\dots,n_d$  is administered, and let  $\omega_{dj}=1-\text{Imm}_{dj}$  be the corresponding objective function coefficient for decision variable  $\pi_{dj}$ . These objective function coefficients place more emphasis on the earlier vaccine dose requirements for each disease. For example, for disease  $d \in D$ , suppose  $\text{Imm}_{d1}=.4$  (after dose  $j=1$ ),  $\text{Imm}_{d2}=.75$  (after dose  $j=2$ ), and  $\text{Imm}_{d3}=0.95$  (after dose  $j=3$ ), and hence,  $\omega_{d1}=.6$ ,  $\omega_{d2}=.25$ , and  $\omega_{d3}=0.05$ . Therefore, satisfying dose  $j=1$  for disease  $d \in D$  is much more attractive to the IP than satisfying dose  $j=3$  for disease  $d \in D$ , since IP is a maximization problem. Furthermore, these objective function coefficients assume that the earlier doses in the childhood immunization schedule are the most important (i.e., it is more important that a child receives the first dose of some vaccine that immunizes against disease  $d \in D$  (e.g., Hepatitis A) than to receive the fourth dose of some vaccine that immunizes against some other disease  $d' \in D$  (e.g., Polio)). This extension of IP assumes equivalent immunogenicities for each vaccine  $v \in V$  that immunizes against disease  $d \in D$ . Weighting the decision variables accordingly would require the additional constraint  $\pi_{d1} \geq \pi_{d2} \geq \dots \geq \pi_{dn_d}$  for all diseases  $d \in D$  in the formulation of IP, which requires, for a given disease  $d \in D$ , that a vaccine for dose  $i$  must be administered before a vaccine for dose  $j$  may be administered for all doses  $i < j$ . The additional constraint  $\pi_{d1} \geq \pi_{d2} \geq \dots \geq \pi_{dn_d}$  for all diseases  $d \in D$  is very restrictive; however, maximizing immunogenicity in a less restrictive manner will require that the constraints in the IP model be modified.

An additional extension to the model IP is to define a lower bound  $l_d$  for each disease  $d \in D$ , where  $l_d$  is a lower bound on the number of doses for disease  $d \in D$  that must be satisfied. Therefore, the additional constraint  $\sum_{j=1}^{n_d} \pi_{dj} \geq l_d$  for all

diseases  $d \in D$  would enforce a lower bound on effectiveness while also prohibiting extraimmunization. Nevertheless, the inclusion of these constraints would likely yield an infeasible model for unreasonable values of  $l_d$ . Theoretically, it is easy to construct an instance of IP that is infeasible for any lower bound greater than zero for a given disease  $d \in D$  (or subset of diseases); however, for all practical instances of IP, the given vaccine formulary will likely allow some doses to be satisfied for all diseases  $d \in D$ , since combination vaccines are designed to immunize against diseases that share common dose requirements in the Recommended Childhood Immunization Schedule (e.g., compare pneumococcus and *Haemophilus influenzae* type b in Fig. 1).

### 3 A Dynamic programming algorithm and randomized rounding heuristic

This section presents a dynamic programming (DP) algorithm that finds an optimal solution for IP. The randomized rounding heuristic for IP is also presented. The randomized rounding heuristic is shown to be a randomized approximation algorithm, which by definition, executes in polynomial time and provides an approximation bound on the expected value of the heuristic solution [17].

#### 3.1 Dynamic programming algorithm

The stages of the DP correspond to the time periods in  $T$ . A state, say  $S^t$ , in stage  $t \in T$  is a vector of length  $\delta$ , where  $S_d^t$  is the number of doses desired to be given for disease  $d \in D$  through stage  $t \in T$ . Let  $m_{dt}$  ( $M_{dt}$ ) be the minimum (maximum) number of doses scheduled for disease  $d \in D$  through stage  $t \in T$ . For example, for the first disease in the Recommended Childhood Immunization Schedule (see Fig. 1),  $m_1=(1, 1, 2, 2, 2, 2, 2, 3, 3, 3)$  and  $M_1=(1, 2, 2, 2, 3, 3, 3, 3, 3, 3)$ . The desired number of doses  $S_d^t$  should satisfy  $m_{dt} \leq S_d^t \leq M_{dt}$ , and hence, the set of states for stage  $t \in T$  is  $\Omega_t = \{S^t \in Z^\delta : m_{dt} \leq S_d^t \leq M_{dt} \text{ for all } d \in D\}$ , where  $Z$  denotes the set of integers. For example, for the first four diseases in Fig. 1, the set of possible states in stage 5 are  $(2, 3, 3, 2)$ ,  $(2, 3, 3, 3)$ ,  $(3, 3, 3, 2)$ ,  $(3, 3, 3, 3)$ .

$S_d^t$  is the number of doses desired to be given for disease  $d \in D$  through stage  $t \in T$ . However, it may not be possible to administer all of the desired doses in a state due to the restriction on extraimmunization. Let  $Z(S^t)$  be the maximum number of desired doses in  $S^t$  that can be administered through stage  $t \in T$  without extraimmunization for any disease  $d \in D_{NE}$ .

Moving from state  $S^{t-1}$  in stage  $t-1$  to state  $S^t$  in stage  $t$  corresponds to administering vaccines in time period  $t \in T$ . If vaccination is permitted, it is desired to administer a vaccine for each disease  $d \in D$  such that  $S_d^t > S_d^{t-1}$ . Note that  $S_d^t \leq$



$S_d^{t-1} + 1$  because the doses for disease  $d \in D$  are mutually exclusive. Furthermore, note that it is infeasible to move from state  $S^{t-1}$  in stage  $t-1$  to state  $S^t$  in stage  $t$  if  $S_d^t < S_d^{t-1}$  for any disease  $d \in D$ , since it is impossible to reduce the number of doses that have been given for disease  $d \in D$ . Therefore, state  $S^{t-1}$  in stage  $t-1$  is connected to state  $S^t$  in stage  $t$  if and only if  $b = S^t - S^{t-1}$  is a binary vector, and hence, for a given stage  $t \in T$ , it is desired to administer vaccines for the set of diseases  $D' = \{d \in D : b_d = S_d^t - S_d^{t-1} = 1\}$ , where  $b_d$  is the  $d$ th component of the binary vector  $b$ . The set of vaccines that may be administered for the diseases in  $D'$  is  $V' = \{v \in V : I_{vd} = 1 \text{ for some } d \in D' \text{ and } I_{vd} = 0 \text{ for all } d \in D_{NE} \setminus D'\}$ . Once  $D'$  and  $V'$  are known, the problem is to find a subset of  $V'$  that protects against as many diseases in  $D'$  as possible without incurring extrimmunization for any disease  $d \in D_{NE}$ . This problem has the same structure as IP, except that it contains only a single time period, and hence, it is called the One Period Problem (OPP( $b$ )).

Define  $g(b)$  to be the optimal value of OPP( $b$ ) and recall that  $Z(S^t)$  is the maximum number of desired doses in  $S^t$  that can be given without extrimmunization. If  $Z(S^t)$  is obtained by moving from  $S^{t-1}$ , then  $Z(S^t) = Z(S^{t-1}) + g(b)$ , where  $b = S^t - S^{t-1}$ . Since it is desired to maximize  $Z(S^t)$  over all states from the previous stage that are connected to  $S^t$ , the dynamic programming recurrence relation is

$$Z(S^t) = \max_{S^{t-1} \in \Omega_{t-1} : b = S^t - S^{t-1} \text{ is binary}} Z(S^{t-1}) + g(b).$$

The final stage contains a single state, namely  $S^\tau = (n_1, n_2, \dots, n_\delta)$ , thus the optimal value is  $Z(S^\tau)$ .

**Theorem 1** *The DP finds an optimal solution for IP.*

*Proof* This result follows by demonstrating that the optimal solution to IP is embedded in the state space of the DP, which is shown by demonstrating that the state space for stage  $t$  is sufficiently connected to the state space in stage  $t-1$  ( $t=2,3,\dots,\tau$ ) and that  $Z(S^\tau)$  equals the optimal value of IP. See Appendix 1 for a detailed proof.  $\square$

Note that OPP( $b$ ) must be solved each time the recurrence relation is used in the DP; however, for a given stage  $t \in T$ , the binary vector  $b$ , which denotes the set of diseases for which a vaccine dose is desired, may have already been encountered in a previous stage, and hence, there is no need to resolve OPP( $b$ ) if the solution has been stored and is available. A “branch and remember” algorithm was used in the computational experiments reported in Section 4 (see the Appendix 1 for the pseudo-code for this algorithm.)

The DP algorithm offers several theoretically and computationally advantages. First, the DP algorithm is efficient in practice with the current Recommended Childhood Immunization Schedule, since this schedule yields a reasonable state/decision space, and the OPP( $b$ ) instances in each stage  $t \in T$

are small. Second, the structure of the DP algorithm is ideal for solving partial (or incomplete) childhood immunization schedules that arise when children have been partially immunized and then re-enter the healthcare system to complete their immunization schedules (where the set of vaccines used to partially immunize the children are feasible initial solutions for the DP algorithm; this problem is termed the *schedule completion problem*. Note that if this set of vaccines is not feasible, then the recommended catch-up schedule would need to be used, which would require a completely new model formulation and analysis; see [5] for details of this catch-up schedule). Lastly, the structure of the DP algorithm makes it easier to capture restrictions that are schedule-specific by imposing such restrictions on each OPP ( $b$ ) instance (see [5, 16]).

### 3.2 Randomized rounding heuristic

The randomized rounding heuristic uses the solution from an LP relaxation to construct a feasible binary solution for IP. This technique has been applied successfully for other discrete optimization problems that share a common structure with IP (e.g., [17, 28]). Relaxing the binary constraints for the decision variables in IP yields the LP relaxation, LP, with objective function ( $O$ ), constraints (1–3), and real decision variables  $0 \leq X_{tv} \leq 1$  for all  $t \in T, v \in V$  and  $0 \leq \pi_{dj} \leq 1$  for all  $d \in D, j=1,2,\dots,n_d$ . Denote the optimal objective function values for IP and LP, as  $z_{IP}$  and  $z_{LP}$ , respectively, where  $z_{LP} \geq z_{IP}$ . Let  $X_{LPtv}^*$  denote the optimal decision vector for LP, with  $X_{LPtv}^*, t \in T, v \in V$ , and  $\pi_{dj}^*, d \in D, j=1,2,\dots,n_d$ , denoting the optimal decision variable values for LP. After solving LP, the randomized rounding heuristic assigns binary decision variable  $X_{tv} = 1(0)$  with probability  $X_{LPtv}^* (1 - X_{LPtv}^*)$  for each time period  $t \in T$  and vaccine  $v \in V$ . Note that constraint (3) in LP implicitly eliminates (i.e., sets to zero) all decision variables  $X_{LPtv}^*$  such that  $R_{dt} = I_{vd} = 1$  for  $d \in D_{NE}$ , and hence, the corresponding binary variable  $X_{tv} = 0$ . This binary variable assignment is then used to determine the number of satisfied doses. The randomized rounding heuristic is now formally given in pseudo-code form.

Randomized rounding heuristic for IP

- Step 1. Solve LP
- Step 2. For  $i=1,2,\dots,K$  (where  $K$  is the user specified number of replications)
  - a.  $X_{tv} \leftarrow 0$  for all  $t \in T$  and  $v \in V$
  - b.  $\pi_{dj} \leftarrow 0$  for all  $d \in D$  and  $j=1,2,\dots,n_d$
  - c. For all  $t \in T$  and  $v \in V$ 
    - i. Draw a random number  $RAND \sim U(0,1)$
    - ii. If  $X_{LPtv}^* \geq RAND$ , then  $X_{tv} \leftarrow 1$

- d. For all  $d \in D_{NE}$ 
    - i. For all  $j=1,2,\dots,n_d$ 
      - 1.  $\hat{\pi}_{dj} \leftarrow \sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd}$
      - 2. If  $\hat{\pi}_{dj} > 1$ , then set  $X_{tv} \leftarrow 0$  such that  $X_{tv}=1$  and  $P_{djt}=I_{vd}=1$  for the ( $\hat{\pi}_{dj} = 1$ ) variables with corresponding smallest fractional values  $X_{LPtv}^*$
  - e. For all  $d \in D$ 
    - i. For all  $j=1,2,\dots,n_d$ 
      - 1. Compute  $\hat{\pi}_{dj} \leftarrow \sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd}$ 
        - a. If  $d \in D_{NE}$  and  $\hat{\pi}_{dj} = 1$ , then  $\pi_{dj} \leftarrow 1$
        - b. If  $d \in D_E$  and  $\hat{\pi}_{dj} > 1$ , then  $\pi_{dj} \leftarrow 1$
  - f. Compute SatisfiedDoses( $i$ ) =  $\sum_{d \in D} \sum_{j=1}^{n_d} \pi_{dj}$  for replication  $i$
- Step 3. Return  $\max_i$  SatisfiedDoses( $i$ )

The randomized rounding heuristic executes in  $O(T_{LP} + K(\tau^2 \nu \delta))$  time, where  $T_{LP}$  is the time required to solve LP. Steps 2.d and 2.e ensure that the solution returned by the randomized rounding heuristic is feasible. Given that linear programming is solvable in polynomial time (assuming an appropriate algorithm; see [1]), it then follows that the randomized rounding heuristic executes in polynomial time. Theorem 2 shows that the expected value of the solution returned by the randomized rounding heuristic is guaranteed to be no worse than  $\xi \times z_{IP}$  where  $\xi = (1 - p_{max})^{\alpha-1}$  for  $p_{max} =$

$$\max_{t \in T, v \in V: 0 < X_{LPtv}^* < 1} \left\{ X_{LPtv}^* \right\} \text{ and } \alpha = \max_{d \in D} (\sum_{v \in V} I_{vd}) \left( \max_{j=1,2,\dots,n_d} \sum_{t \in T} P_{djt} \right),$$

which is the maximum number of non-zero columns in any row of the constraint matrix for IP.

**Theorem 2** *The Randomized Rounding heuristic is a randomized  $\xi$ -approximation algorithm for IP.*

*Proof* See the Proof of Theorem 2 in Appendix 1 for a detailed proof.  $\square$

Theorem 2 shows that, in the worst case, the randomized rounding heuristic could return a solution that is arbitrarily close to zero; however, the computational results in Section 4 show that the randomized rounding heuristic performs well in practice for several randomly generated IP instances.

### 4 Computational results

This section reports computational results comparing the randomized rounding heuristic for IP with the DP algorithm and an IP branch and bound (IP B&B) algorithm that both optimally solve IP. The randomized rounding heuristic and the DP and IP B&B algorithms were executed on three sets

of test problems to demonstrate their computational effectiveness and limitations. The first test problem is the 2006 Recommended Childhood Immunization Schedule. The second set of test problems are randomly generated based on hypothetical near-term future childhood immunization schedules, while the third set of test problems are larger, randomly generated childhood immunization schedules executed with several different vaccine sets. The size of these randomly generated childhood immunization schedules assume that future Recommended Childhood Immunization Schedules will expand to include more diseases and time periods, and hence, will require a larger number of both monovalent and combination vaccines [9, 11, 18].

The solution quality effectiveness measure for the randomized rounding heuristic and the DP and IP B&B algorithms is  $\zeta$ , which is the value of the objective function  $Z$  divided by the total number of required doses, and hence,  $\zeta = Z / \sum_{d=1}^{\delta} n_d$ . Therefore,  $\zeta$  represents the percentage of dose requirements satisfied by the randomized rounding heuristic or the DP and IP B&B algorithms. The execution time (in CPU seconds) is also reported for the heuristic and each exact algorithm, which is the efficiency effectiveness measure. The randomized rounding heuristic and the DP and IP B&B algorithms were coded and executed in *MATLAB*v7.0 on a 2.4 GHz Pentium IV. The IP B&B algorithm used the existing binary solver in the *MATLAB* optimization toolbox.

The first test problem is the 2006 Recommended Childhood Immunization Schedule (see Fig. 1). Therefore,  $D = \{1 = \text{Hepatitis B}, 2 = \text{Diphtheria-Tetanus-Pertussis}, 3 = \text{Haemophilus influenzae type b}, 4 = \text{Polio}, 5 = \text{Measles-Mumps-Rubella}, 6 = \text{Varicella}, 7 = \text{Pneumococcus}, 8 = \text{Influenza}, 9 = \text{Hepatitis A}\}$  with dose vector  $n = (3, 5, 4, 4, 2, 1, 4, 1, 2)$ , since diphtheria, tetanus, and pertussis are considered one disease and measles, mumps, and rubella are also considered one disease, and  $T = \{1, 2, \dots, 10\}$ . The schedule parameters  $P_{djt}$ ,  $Q_{djt}$ , and  $R_{djt}$  for diseases  $d \in D$ , dose  $j = 1, 2, \dots, n_d$ , and time periods  $t \in T$  are all obtained from Fig. 1. For example, for disease  $d = 1 = \text{Hepatitis B}$  and dose  $j = 2$ ,  $P_{djt} = 1(0)$  for time periods  $t = 2, 3(1, 4, 5, 6, 7, 8, 9, 10)$ . Two different vaccine formularies,  $V_1$  and  $V_2$ , are evaluated on two different sets of diseases that restrict extraimmunization,  $D_{NE1}$  and  $D_{NE2}$ . The vaccine formularies are  $V_1 = \{1 = \{1\}, 2 = \{2\}, 3 = \{3\}, 4 = \{4\}, 5 = \{5\}, 6 = \{6\}, 7 = \{7\}, 8 = \{8\}, 9 = \{9\}, 10 = \{2, 3\}, 11 = \{1, 3\}, 12 = \{1, 2, 4\}, 13 = \{5, 6\}\}$  and  $V_2 = \{1 = \{1\}, 2 = \{2, 3, 4\}, 3 = \{1, 9\}, 4 = \{4\}, 5 = \{5\}, 6 = \{6\}, 7 = \{7\}, 8 = \{8\}, 9 = \{9\}, 11 = \{1, 3\}, 12 = \{1, 2, 4\}, 13 = \{5, 6\}, 14 = \{1, 2, 3, 4\}\}$ , where  $V_1$  represents a set of pediatric vaccines currently licensed for use in the United States and  $V_2$  represents a formulary of pediatric vaccines with fewer monovalent vaccines and more combination vaccines, some of which are not yet licensed for use in the United States, but are projected to be in the future. The parameters  $I_{vd}$  are indicated within the sets  $V_1$  and  $V_2$ ,

respectively. For example, vaccine  $1 \in V_1$  is the monovalent vaccine for disease 1 (Hepatitis B), vaccine  $12 \in V_1$  is the combination vaccine *Pediarix*<sup>®</sup> that immunizes against diseases 1 (Hepatitis B), 2 (Diphtheria-Tetanus-Pertussis), and 4 (Polio), and vaccine  $2 \in V_2$  is the combination vaccine *Pentacel*<sup>®</sup> that immunizes against diseases 2 (Diphtheria-Tetanus-Pertussis), 3 (*Haemophilus influenzae* type b), and 4 (Polio), which has been submitted for licensing in the United States. The disease sets are  $D_{NE1} = \{1,2,3,4\}$  and  $D_{NE2} = D = \{1,2,3,4,5,6,7,8,9\}$ . Table 1 reports the solution quality and execution time (in CPU seconds) for the heuristic and the exact algorithms, for each vaccine set and disease set combination. Furthermore,  $K=10$  in the randomized rounding heuristic for each vaccine set and disease set combination.

IP is solvable in polynomial time when all the vaccines in  $V$  are monovalent vaccines or when there exists a corresponding monovalent vaccine for every disease in  $D$ , and hence, the results for the solution quality and execution time reported in Table 1 are not surprising, given that most diseases have a corresponding monovalent vaccine, particularly in vaccine set  $V_1$  [16]. Moreover, the randomized rounding heuristic returned the optimal solution for each case, but, in most cases, took more time to execute than the exact algorithms since the random rounding was replicated  $K=10$  times. However, as the next set of test problems will illustrate, this is unlikely to occur for future Recommended Childhood Immunization Schedules, as the schedule expands and more combination vaccines gain FDA approval.

The second set of test problems considers hypothetical near-term future childhood immunization schedules. The randomized rounding heuristic and the DP and IP B&B algorithms were executed on 30 randomly generated childhood immunization schedules with 15 time periods, 11 diseases, and a vaccine formulary of 30 vaccines. Therefore, each random childhood immunization schedule reflects a gradual expansion in the sets  $D$  (from 9 to 11 diseases) and  $T$  (from 10 to 15 time periods) and a modest increase in the number of available vaccines, particularly, combination vaccines. In each random childhood immunization schedule,  $1 \leq n_d \leq 5$  for all diseases  $d \in D$ ,  $1 \leq Val(v) \leq 6$  for all vaccines  $v \in V$ , and  $P_{djt} = 1$  for at most three time periods  $t \in T$  for every disease  $d \in D$  and dose  $j = 1, 2, \dots, n_d$ . For each randomly generated childhood immunization

schedule, the randomized rounding heuristic and both exact algorithms were executed three times, where in execution 1, 2, and 3,  $d_{NE} = 4, 8,$  and  $11$ , respectively. Furthermore, in each execution, the randomized rounding heuristic was replicated for  $K=100, 300,$  and  $500$ . Table 2 reports the solution quality and execution time (in CPU seconds) averaged over the 30 random childhood immunization schedules for each value of  $d_{NE}$  (and  $K$  for the randomized rounding heuristic). An additional measure  $\lambda$ , which indicates the number of childhood immunization schedules that the respective heuristic or algorithm found the optimal solution, is also reported. For the IP B&B algorithm, the optimal solution was not always found because the algorithm exceeded the default execution time limit (2 h) or default iteration limit ( $10^7$ ). Therefore, the statistics reported in Table 2 for the IP B&B algorithm are averaged over the  $\lambda$  random childhood immunization schedules for which the algorithm found the optimal solution.

The data reported in Table 2 show that the randomized rounding heuristic found better solutions when  $\delta_{NE} \ll \delta$ , which is consistent with Theorem 2. Recall from the proof of Theorem 2 that it can be shown that the bound on the expected solution value returned by the randomized rounding heuristic improves when  $D_{NE} = \emptyset$  (i.e.,  $\delta_{NE} = 0$ ). Moreover, the randomized rounding heuristic, on average, was always more efficient than the IP B&B algorithm. Furthermore, the IP B&B algorithm’s ability to find an optimal solution in a ‘reasonable’ amount of time and memory was sensitive to the value of  $d_{NE}$ . Moreover, the randomized rounding heuristic was more efficient than the DP algorithm for  $\delta_{NE} = 8$  and  $11$ , provided  $K=100$ . The DP algorithm was both efficient (outperforming the randomized rounding heuristic for  $K=300$  and  $500$  and the IP B&B algorithm in all cases) and effective (found an optimal solution in each execution).

The observed difference in execution time between the randomized rounding heuristic or DP algorithm and the IP B&B algorithm reported in Table 2 could become problematic for larger childhood immunization schedules and/or for practical uses. For example, a webpage used to find a vaccine formulary for a given childhood immunization schedule that prohibits extraimmunization would require an algorithm to execute in real-time, since most web users would terminate a web application that required several

**Table 1** Computational results for 2006 recommend childhood immunization schedule

Algorithm	$V_1$ and $D_{NE1}$		$V_1$ and $D_{NE2}$		$V_2$ and $D_{NE1}$		$V_2$ and $D_{NE2}$	
	$\zeta$	Time	$\zeta$	Time	$\zeta$	Time	$\zeta$	Time
randomized rounding	1.00	0.15	1.00	0.54	0.96	0.56	0.96	0.57
DP	1.00	0.27	1.00	0.26	0.96	0.26	0.96	0.33
IP B&B	1.00	0.20	1.00	0.27	0.96	0.22	0.96	0.22



**Table 2** Computational results for future childhood immunization schedule

Algorithm	$d_{NE}=4$			$d_{NE}=8$			$d_{NE}=11$		
	$\zeta$	Time	$\lambda$	$\zeta$	Time	$\lambda$	$\zeta$	Time	$\lambda$
Randomized, $K=100$	0.972	0.36	24	0.722	0.44	8	0.549	0.38	5
Randomized, $K=300$	0.975	0.54	25	0.727	0.67	8	0.564	0.66	10
Randomized, $K=500$	0.976	0.72	25	0.729	0.93	8	0.569	0.93	13
DP	0.98	0.24	30	0.77	0.50	30	0.59	0.61	30
IP B&B	0.98	1.06	30	0.77	3.91	29	0.59	7.48	25

seconds or minutes to execute. Moreover, the difference in execution time between the randomized rounding heuristic or DP algorithm and the IP B&B algorithm will provide a more efficient analysis of larger childhood immunization schedules that may involve Monte Carlo simulation (see [22]), where either of these may require the solution of hundreds of thousands of integer programming instances. Furthermore, the childhood immunization schedule may need to be solved for each child, on a case-by-case basis, depending on the desired diseases in the set  $D_{NE}$ , and hence, efficient algorithms are needed so as to provide, in real-time, practical value for the public health community.

The third set of test problems considers larger randomly generated childhood immunization schedules that demonstrate the effect of combination vaccines and further demonstrate how a childhood immunization schedule’s size affects the efficiency and solution quality of the randomized rounding heuristic and the exact algorithms. The randomized rounding heuristic and the exact algorithms were executed on 30 randomly generated childhood immunization schedules with 20 time periods, 13 diseases, and a vaccine formulary of 50 vaccines, where  $\delta_{NE}=10$ ,  $1 \leq n_d \leq 5$  for all diseases  $d \in D$ , and  $P_{djt}=1$  for at most four time periods  $t \in T$  for every disease  $d \in D$  and dose  $j=1,2,\dots,n_d$ . For each randomly generated childhood immunization schedule, the randomized rounding heuristic and both exact algorithms were executed six times, where for execution  $i=1,2,\dots,6$ ,  $Val$

$(v) \leq i$  for all vaccines  $v \in V$ . Table 3 reports the solution quality and execution time (in CPU seconds) averaged across all 30 randomly generated childhood immunization schedules as well as the measure  $\lambda$  for each valency. The overall average and standard deviation across all vaccine sets is also reported. Note that the IP B&B algorithm found the optimal solution for 151 of the 180 instances (each random childhood immunization schedule was executed six times), but exceeded the maximum number of iterations allowed for the remaining 29 instances (three instances for  $Val(v) \leq 4$ , 11 instances for  $Val(v) \leq 5$ , and 15 instances for  $Val(v) \leq 6$ ). The statistics reported in Table 3 for the IP B&B algorithm are for the 151 instances for which the IP B&B algorithm found the optimal solution while the execution time in parenthesis is the average over all instances. Furthermore,  $K=300$  in each execution of the randomized rounding heuristic.

The data reported in Table 3 demonstrate how the size of the childhood immunization schedule and the complexity of the vaccine set impact the execution time and solution quality of the randomized rounding heuristic and the exact algorithms. For example, in most cases, the execution time required to execute the heuristic and exact algorithms steadily increased as the valency (i.e., complexity) of the vaccine formulary increased. As expected, the solution quality of the randomized rounding heuristic deteriorated as the valency of the vaccine formulary increased, since as the valency of the vaccine formulary increases, the number of fractional varia-

**Table 3** Computational Results for the Effect of Combination Vaccines

Val( $v$ ) $\leq$	Algorithm								
	Randomized rounding			DP			IP B&B		
	$\zeta$	Time	$\lambda$	$\zeta$	Time	$\lambda$	$\zeta$	Time	$\lambda$
1	1.00	0.97	30	1.00	1.1	30	1.00	1.1	30
2	0.99	0.98	24	0.99	2.1	30	0.99	1.3	30
3	0.92	1.1	5	0.94	2.5	30	0.94	2.7	30
4	0.87	1.1	2	0.92	5.1	30	0.92	8.6 (1184)	27
5	0.79	1.2	1	0.87	5.3	30	0.87	27.5 (4210)	19
6	0.71	1.2	0	0.77	7.4	30	0.77	8.9 (5957)	15
Average	0.88	1.1	14	0.91	3.9	30	0.91	8.4 (1893)	25
SD	0.11	0.1	12.8	0.09	2.4	0	0.09	10.0 (2574)	6.6

bles in LP should increase (since each decision variable is present in more constraints), which results in a larger gap between the LP optimal solution value and the IP optimal solution value. Therefore, since the randomized rounding heuristic solution is derived from the LP optimal solution, the gap between the randomized rounding heuristic solution value and IP optimal solution value also increases. Theorem 2 offers theoretical support for this claim, since  $p_{\max}$  is likely to be larger as the valency of the vaccine formulary increases. However, the execution time of the randomized rounding heuristic was insensitive to the valency of the vaccine formulary versus the exact algorithms, which were more sensitive to the valency of the vaccine formulary. Furthermore, the randomized rounding heuristic was more efficient than both exact algorithms; however, the DP algorithm was twice as fast as the IP B&B algorithm on average for the instances that the IP B&B algorithm was able to optimally solve and was nearly 500 times as fast as the IP B&B algorithm when averaged across all 180 instances. Moreover, the DP algorithm effectively solved all 180 instances to optimality.

## 5 Conclusion and research extensions

This paper reports results on applying operations research techniques to further examine the issue of extraimmunization in pediatric immunization by extending the results in Hall [16]. There are several articles in the literature that report research that applies operations research techniques to pediatric immunization problems (e.g., [20–22, 24–26]). This paper moves in a new direction by applying discrete optimization techniques to address the issue of pediatric vaccine extraimmunization. A discrete optimization model that seeks to maximize the effectiveness of a vaccine formulary while prohibiting extraimmunization for a given childhood immunization schedule was formulated in Section 2. As more combination vaccines come to market and the Recommended Childhood Immunization Schedule becomes more complex to include more diseases and cover more time periods, this discrete optimization model will capture the combinatorial explosion of alternatives for public health policy-makers and administrators, vaccine manufacturers, pediatricians, and parents/guardians by identifying vaccine formularies and schedules that safely use combination vaccines, which will help address safety concerns in pediatric immunization, reduce costs, and reduce vaccine wastage associated with extraimmunization.

An exact DP algorithm as well as a randomized rounding heuristic were presented in Section 3 and compared computationally to an IP B&B algorithm in Section 4. These results showed that for larger randomly generated childhood immunization schedules, the randomized rounding heuristic

was more efficient than the IP B&B algorithm and provided a reasonable solution quality that, on average, was within 5% of the optimal solution value. Furthermore, the randomized rounding heuristic was able to find a feasible solution in seconds for all random childhood immunization schedules, whereas the IP B&B algorithm often exceeded default time and memory limits when  $\max_{v \in V} Val(v) \geq 4$ . Moreover, the results showed the DP algorithm to be significantly more efficient and effective than the IP B&B algorithm. Therefore, for practical purposes, the results suggest that the DP algorithm is the algorithm of choice, since it is an efficient exact algorithm. However, with the significant expansion of the Recommended Childhood Immunization Schedule over the past decade (see [5, 9]) and with numerous new pediatric vaccines being considered (see [18]), the randomized rounding heuristic will allow more efficient analysis of larger childhood immunization schedules and practical analysis involving Monte Carlo simulation or finding an optimal vaccine formulary for each child on a case-by-case basis, which will require the solution of several unique integer programming instances. Advances in biotechnology and differing perspectives on the safety of combination vaccines will only heighten the possibility for considering each child on a case-by-case basis (see [19]).

Work is in progress to determine the trade-off of satisfying a given childhood immunization schedule while minimizing extraimmunization versus strictly prohibiting extraimmunization and maximizing the number of vaccine doses satisfied, since both of these alternatives incur different costs. The model presented in this paper will allow such trade-offs to be analyzed. Moreover, further examining the model extensions (theoretically and empirically) presented in Section 2 is an area of future research since these extensions offer practical value but may also significantly alter the structure of the model, which would require entirely different solution techniques. Finally, work is in progress to further extend the results reported by determining new heuristics for IP that either empirically provide better solutions and/or improve the existing approximation bound in Section 3.

**Acknowledgements** The authors would like to thank the three anonymous referees for their excellent comments and input, which have resulted in a significantly improved manuscript. The authors would also like to thank Bruce G. Weniger, M.D., M.P.H., Assistant Chief for Vaccine Development, Vaccine Safety and Development Branch, National Immunization Program, Centers for Disease Control and Prevention, for his longstanding encouragement and feedback on this line of research; his input on this work has been invaluable and most appreciated. This research has been supported in part by the National Science Foundation (DMI-0457176, DMI-0456945). The third author has also been supported in part by the Air Force Office of Scientific Research (FA9550-07-1-0232). The views expressed in this paper are those of the authors and do not reflect the official policy or

position of the United States Air Force, Department of Defense, National Science Foundation, or the United States Government.

### Appendix 1

#### Proof of Theorem 1

Let  $X^*, \pi^*$  be an optimal solution of IP. The proof proceeds by demonstrating that a solution corresponding to  $X^*, \pi^*$  is embedded in the state space of DP. The proof requires a few additional definitions. Let  $T_{dj} = \{t \in T: P_{djt} = 1\}$  to be the set of time periods when vaccine dose  $j$  ( $=1, 2, \dots, n_d$ ) may be administered for disease  $d \in D$ .  $T_{dj}$  is called the  $j$ th time window for disease  $d \in D$ . Given a disease  $d \in D$  and a time period  $t \in T$ , define  $\text{window}(d, t)$  to be

$$\text{window}(d, t) = \begin{cases} T_{dj} & \text{if } t \in T_{dj} \\ \emptyset & \text{if } t \notin \cup_{j=1}^{n_d} T_{dj}, \end{cases}$$

which is the time window for disease  $d$  that contains time period  $t$  (recall that the doses for a given disease are mutually exclusive, which implies that  $t$  can belong to at most one time window for a given disease). Define

$$y_{dt} = \begin{cases} 1 & \text{if there exists } v, j \text{ such that } X_{tv}^* = 1, \\ & P_{djt} = 1, X_{tv}^* = 0, \text{ for all } t' \in T_{dj} \text{ with } t' < t \\ & \text{and } I_{vd} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the variable  $y_{dt}$  equals one if  $t$  is in a time window for disease  $d$ , a vaccine that protects against disease  $d$  was administered in time period  $t$ , and no vaccine for disease  $d$  was administered in an earlier time period in the same time window as  $\text{window}(d, t)$ .

Let

$$S_d^t = \begin{cases} M_{dt} & \text{if } y_{dt} = 1 \text{ for some } t' \text{ such that } t' \leq t \text{ and} \\ & t' \in \text{window}(d, t) \\ m_{dt} & \text{otherwise.} \end{cases}$$

Hence,  $S_d^t = M_{dt}$ , if a dose for disease  $d$  was given in time period  $t$  or a previous time period in  $\text{window}(d, t)$ . By definition,  $S^t$  is a state in DP for  $t \in T$ . The remainder of the proof demonstrates that  $S^{t-1}$  is connected to  $S^t$  for  $t=2, 3, \dots, t$  and that  $Z(S^t)$  equals the optimal value of IP, which is  $\sum_{d \in D} \sum_{j=1}^{n_d} \pi_{dj}^*$ .

Recall that state  $S^{t-1}$  is connected to state  $S^t$  if  $b = S^t - S^{t-1}$  is a binary vector. This will be demonstrated by proving that

$$S_d^t = \begin{cases} S_d^{t-1} + 1 & \text{if } y_{dt} = 1 \text{ or } S_d^{t-1} = m_{d,t-1} < m_{dt} \\ S_d^{t-1} & \text{otherwise.} \end{cases}$$

Given that the doses are sequentially ordered, mutually exclusive, and consecutive, it follows from the definitions of  $m_{dt}$  and  $M_{dt}$  that  $M_{dt} = m_{dt}$  if  $t$  is not in a time window for disease  $d$  or if  $t$  is the final time period in a time window for disease  $d$ , otherwise  $M_{dt} = m_{dt} + 1$ . Furthermore,  $m_{dt} = m_{d,t-1}$  unless  $t$  is the final time period in a time window for disease  $d$ , in which case  $m_{dt} = m_{d,t-1} + 1$ . Combining these facts yields  $M_{dt} \leq m_{d,t-1} + 1$ . Moreover,  $S_d^t \leq M_{dt}$  and  $S_d^{t-1} \geq m_{d,t-1}$ , by the definition above, and hence,  $S_d^t \leq M_{dt} \leq m_{d,t-1} + 1 \leq S_d^{t-1} + 1$ .

Suppose that  $y_{dt} = 1$ . Then  $S_d^t = M_{dt}$  and a vaccine was administered for disease  $d$  in time period  $t$ , but no other vaccine for disease  $d$  was administered earlier in  $\text{window}(d, t)$ . Either  $t-1 \in \text{window}(d, t)$ , in which case  $S_d^{t-1} = m_{d,t-1}$  and  $m_{d,t-1} < M_{dt}$ , which in turn implies  $S_d^{t-1} < S_d^t$ , or  $t-1 \notin \text{window}(d, t)$ , in which case  $M_{d,t-1} < M_{dt}$ , which in turn implies  $S_d^{t-1} < S_d^t$ . In either case,  $S_d^{t-1} < S_d^t$  implies  $S_d^t = S_d^{t-1} + 1$ .

Now suppose that  $S_d^{t-1} = m_{d,t-1} < m_{dt}$ . Then  $S_d^{t-1} < S_d^t$ , and hence,  $S_d^t = S_d^{t-1} + 1$ .

Now suppose that  $y_{dt} = 0$  and it is not the case that  $S_d^{t-1} = m_{d,t-1} < m_{dt}$ . If  $m_{dt} < S_d^t$ , then  $S_d^t = M_{dt}$  and a vaccine for disease  $d$  must have been administered in  $\text{window}(d, t)$ . But  $y_{dt} = 0$  implies that a vaccine for disease  $d$  must have been administered in an earlier time period in  $\text{window}(d, t)$ . Therefore,  $S_d^{t-1} = M_{d,t-1} = M_{dt} = S_d^t$ . If  $m_{d,t} = S_d^t$ , then the only scenario under which  $S_d^t = S_d^{t-1} + 1$  occurs is when  $S_d^{t-1} = m_{d,t-1} < m_{dt}$ . Furthermore,  $S_d^{t-1} > S_d^t$  cannot occur because  $M_{d,t-1}$  exceeds  $m_{dt}$  only when  $t-1$  and  $t$  are in the same time window. By construction, if  $S_d^{t-1} = M_{d,t-1}$  and  $t-1$  and  $t$  are in the same time window, then  $S_d^t = M_{dt}$ . Consequently, the suppositions of this paragraph ensure that  $S_d^{t-1} = S_d^t$ . This completes the proof that state  $S^{t-1}$  is connected to state  $S^t$ .

It remains to show that  $Z(S^t)$  equals the optimal value of IP. Let  $b = S^t - S^{t-1}$ . Then

$$b_d = \begin{cases} 1 & \text{if } y_{dt} = 1 \text{ or } S_d^{t-1} = m_{d,t-1} < m_{dt} \\ 0 & \text{otherwise.} \end{cases}$$

Define  $D' = \{d \in D: b_d = 1\}$  and  $V' = \{v \in V: I_{vd} = 1 \text{ for some } d \in D' \text{ and } I_{vd} = 0 \text{ for every restricted } d \in D_{NE} \setminus D'\}$ . Let  $V^* = \{v \in V: X_{tv}^* = 1 \text{ and } v \text{ contains an antigen for a disease } d \text{ such that } y_{dt} = 1\}$ . Suppose  $v \in V^*$ . Then  $v$  contains an antigen for a disease  $d$  such that  $y_{dt} = 1$ , so  $v$  contains an antigen for at least one  $d \in D'$ . If  $v$  contains an antigen for a restricted disease  $d$ , then  $\pi_{dj}^* = 1$  for some  $j=1, 2, \dots, n_d$  such that  $t$  is in the time window for dose  $j$ . Hence  $\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} = 1$ , which implies that  $y_{dt} = 1$ , thus  $d \in D'$ . Consequently,  $v \in V'$  and  $V^* \subseteq V'$ . Therefore,  $V'$  contains a set of vaccines that can be administered in time period  $t$  and contain antigens for at least  $\sum_{d \in D} y_{dt}$  diseases. By the DP

recurrence relation,  $Z(S^t) \geq Z(S^{t-1}) + \sum_{d \in D} y_{dt}$ . Applying this result for  $t=2, 3, \dots, \tau$  yields

$$\begin{aligned} Z(S^\tau) &\geq Z(S^{\tau-1}) + \sum_{d \in D} y_{d\tau} \\ &\geq Z(S^{\tau-2}) + \sum_{d \in D} y_{d,\tau-1} + \sum_{d \in D} y_{d\tau} \\ &\vdots \\ &\geq \sum_{t \in T} \sum_{d \in D} y_{dt}. \end{aligned}$$

Finally,  $\sum_{d \in D} \sum_{j=1}^{n_d} \pi_{dj}^* \leq \sum_{t \in T} \sum_{d \in D} y_{dt}$ , hence the optimal value of IP is less than or equal to  $Z(S^\tau)$ .  $\square$

Branch and Remember Algorithm Pseudo-Code

The following “branch and remember” algorithm was used in the computational experiments reported in Section 4 for the dynamic programming recurrence relation. This algorithm not only stores the solution for  $OPP(b)$  at the root problem, but also stores the solution for all the subproblems that are generated.

$OPP(b)$

if  $OPP(b)$  has been solved previously, then return its optimal value

$D' = \{d \in D: b_{d'} = 1\}$

if  $D' = \emptyset$ , then return(0)

$V' = \{v \in V: I_{vd} = 1 \text{ for some } d \in D' \text{ and } I_{vd} = 0 \text{ for every restricted } d \in D_{NE} \setminus D'\}$

Choose  $d' \in D'$

$max\_doses = 0$

for each  $v \in V'$  such that  $I_{vd'} = 1$ ,

$$b'_{d'} = \begin{cases} 1 & \text{if } b_{d'} = 1 \text{ and } I_{vd'} = 0 \\ 0 & \text{otherwise} \end{cases}$$

$n\_covered = |\{d' \in D': I_{vd'} = 1\}|$

$doses = OPP(b')$

if  $doses + n\_covered > max\_doses$ , then  $max\_doses = doses + n\_covered$

endfor

$$b'_{d'} = \begin{cases} b_{d'} & \text{if } d' \neq d \\ 0 & \text{otherwise} \end{cases}$$

$doses = OPP(b')$

if  $doses > max\_doses$ , then  $max\_doses = doses$

Store  $max\_doses$  for  $b$

return( $max\_doses$ )

Proof of Theorem 2

A randomized approximation algorithm is an algorithm that executes in polynomial time and guarantees the expected

value of its returned solution is within some constant factor  $\xi$  (i.e., randomized rounding is a randomized  $\xi$ -approximation algorithm if  $E[z] \geq \xi \cdot z_{IP}$ , where  $z$  is the value of the objective function returned by the randomized rounding heuristic and  $z_{IP}$  is the optimal value for IP. Clearly, the randomized rounding heuristic executes in polynomial time (i.e.,  $O(T_{LP} + K(\tau^2 v \delta))$ ) given that linear programming is solvable in polynomial time. Recall that  $T_{LP}$  is the time required to solve LP. For purposes of the approximation bound, assume  $K=1$  and define the following random variable

$$\hat{\pi}_{dj} = \begin{cases} 1 & \text{if disease } d \in D, \text{ dose } j = 1, 2, \dots, n_d \text{ is satisfied} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,  $\hat{\pi}_{dj} = 1$  if and only if  $\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} \geq 1$  for disease  $d \in D_E$ , dose  $j = 1, 2, \dots, n_d$  and  $\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} = 1$  for disease  $d \in D_{NE}$ , dose  $j = 1, 2, \dots, n_d$ . Furthermore, note the statistical independence of  $\hat{\pi}_{dj}$  for all  $d \in D, j = 1, 2, \dots, n_d$ . Moreover, the objective function value returned by the randomized rounding heuristic is  $z = \sum_{d \in D} \sum_{j=1}^{n_d} \hat{\pi}_{dj}$ , and hence,

$$\begin{aligned} E[z] &= E \left[ \sum_{d \in D} \sum_{j=1}^{n_d} \hat{\pi}_{dj} \right] = \sum_{d \in D} \sum_{j=1}^{n_d} E[\hat{\pi}_{dj}] \\ &= \sum_{d \in D} \sum_{j=1}^{n_d} (0 \cdot P[\hat{\pi}_{dj} = 0] + 1 \cdot P[\hat{\pi}_{dj} = 1]) \\ &= \sum_{d \in D} \times \sum_{j=1}^{n_d} P[\hat{\pi}_{dj} = 1]. \end{aligned}$$

Observe that  $P[\hat{\pi}_{dj} = 1] = P[\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} \geq 1]$  for disease  $d \in D_E$ , dose  $j = 1, 2, \dots, n_d$  and  $P[\hat{\pi}_{dj} = 1] = P[\sum_{t \in T} \sum_{v \in V} P_{djt} X_{tv} I_{vd} = 1]$  for disease  $d \in D_{NE}$ , dose  $j = 1, 2, \dots, n_d$ . Therefore, for disease  $d \in D_E$ , dose  $j = 1, 2, \dots, n_d$ ,

$$P[\hat{\pi}_{dj} = 1] = P[\text{at least one } X_{tv} = 1] = 1 - P[\text{all } X_{tv} = 0],$$

and for disease  $d \in D_{NE}$ , dose  $j = 1, 2, \dots, n_d$ ,

$$\begin{aligned} P[\hat{\pi}_{dj} = 1] &= P[\text{exactly one } X_{tv} = 1] \\ &= 1 - (P[\text{more than one } X_{tv} = 1] + P[\text{all } X_{tv} = 0]). \end{aligned}$$

Furthermore,  $P[\text{at least one } X_{tv} = 1] \geq P[\text{exactly one } X_{tv} = 1]$  for all diseases  $d \in D$  and doses  $j = 1, 2, \dots, n_d$ , and hence, for purposes of the approximation bound, assume that  $D_E = \emptyset$ . (It can be shown that the randomized rounding heuristic is a randomized  $(1-1/e)$ -approximation algorithm for IP when  $D_{NE} = \emptyset$  by using similar arguments from the Max-Satisfiability Problem [17].)

Consider some disease  $d \in D$  and dose  $j = 1, 2, \dots, n_d$ , then the LP relaxation of IP implies  $\pi_{dj}^* = \sum_{t \in T} \sum_{v \in V} P_{djt}$



$X_{LP_{iv}}^* I_{vd}$ . Furthermore, for this constraint, suppose there are  $k$  fractional variables such that  $0 < X_{LP_{iv}}^* < 1$ . (If  $X_{LP_{iv}}^* = 1$  for some  $t \in T$  and  $v \in V$ , then  $P[\hat{\pi}_{dj} = 1] = 1$ , and if  $X_{LP_{iv}}^* = 0$  for all  $t \in T$  and  $v \in V$ , then  $P[\hat{\pi}_{dj} = 1] = 0$ ). Then,

$$P[\hat{\pi}_{dj} = 1] = P[\text{exactly one } X_{iv} = 1] = X_{LP_{iv_1}}^* (1 - X_{LP_{iv_2}}^*) \dots (1 - X_{LP_{iv_k}}^*) + X_{LP_{iv_2}}^* (1 - X_{LP_{iv_3}}^*) \dots (1 - X_{LP_{iv_k}}^*) + \dots + X_{LP_{iv_k}}^* (1 - X_{LP_{iv_1}}^*) \dots (1 - X_{LP_{iv_{k-1}}}^*)$$

which implies,

$$P[\hat{\pi}_{dj} = 1] \geq X_{LP_{iv_1}}^* (1 - p)^{k-1} + X_{LP_{iv_2}}^* (1 - p)^{k-1} + \dots + X_{LP_{iv_k}}^* (1 - p)^{k-1}, \text{ where } p = \max_{i=1,2,\dots,k} X_{LP_{iv_i}}^* = (1 - p)^{k-1} (X_{LP_{iv_1}}^* + X_{LP_{iv_2}}^* + \dots + X_{LP_{iv_k}}^*) = (1 - p)^{k-1} \pi_{dj}^*.$$

Therefore,

$$E[z] = \sum_{d \in D} \sum_{j=1}^{n_d} P[\hat{\pi}_{dj} = 1] \geq \max_k (1 - p_{\max})^{(k-1)} \sum_{d \in D} \sum_{j=1}^{n_d} \pi_{dj}^*, \text{ where } p_{\max} = \max_{t \in T, v \in V: 0 < X_{LP_{iv}}^* < 1} \{X_{LP_{iv}}^*\} \geq (1 - p_{\max})^{(\alpha-1)} \cdot z_{LP}, \text{ where } \alpha = \max_{d \in D} \alpha_d \text{ and } \alpha_d = (\sum_{v \in V} I_{vd}) \left( \max_{j=1,2,\dots,n_d} \sum_{t \in T} P_{djt} \right) \geq (1 - p_{\max})^{(\alpha-1)} \cdot z_{IP} \text{ (since } z_{LP} \geq z_{IP}), \text{ and hence, } \xi = (1 - p_{\max})^{(\alpha-1)} \text{ for the Randomized Rounding heuristic.}$$

**References**

1. Bazaraa MS, Jarvis JJ, Sherali HD (1990) Linear programming and network flows. Wiley, New York
2. Burden RL, Faires JD (1997) Numerical analysis, 6th edn. Brookes-Cole, New York
3. CDC Vaccine Price List (2006) Center for disease control and prevention, national immunization program website. Retrieved August 3, 2006, from [www.cdc.gov/nip/vfc/cdc\\_vac\\_price\\_list.htm](http://www.cdc.gov/nip/vfc/cdc_vac_price_list.htm).

4. Centers for Disease Control and Prevention (2006) Recommended childhood and adolescent immunization schedule—United States. *Morb Mort Wkly Rep* 54(52):Q1–Q4
5. Centers for Disease Control and Prevention (2007) Recommended childhood and adolescent immunization schedule—United States. *Morb Mort Wkly Rep* 55(51):Q1–Q4
6. Centers for Disease Control and Prevention (2003) Prevention and control of infections with hepatitis viruses in correctional settings. Appendix hepatitis A and B vaccines. *Morb Mort Wkly Rep* 52(RR-1):34–36
7. Centers for Disease Control and Prevention (2002) General recommendations on immunization. *Morb Mort Wkly Rep* 51(RR-2)
8. Centers for Disease Control and Prevention (1999) Combination vaccines for childhood immunization. *Morb Mort Wkly Rep* 48(18)(RR-5)
9. Centers for Disease Control and Prevention (1995) Recommended childhood and adolescent immunization schedule—United States. *Morb Mort Wkly Rep* 44(RR-5):1–9
10. Chen RT, DeStefano F, Pless R, Mootrey G, Kramarz P, Hibbs B (2001) Challenges and controversies in immunization safety. *Infect Dis Clin North Am* 15(1):21–39
11. Cochi SL (2005) 2005 National immunization survey. Presentation to National Press Club, July 27, 2005
12. Diekema DS (2005) Responding to parental refusals of immunization of children. *Pediatrics* 115(5):1428–1431
13. Edwards KM, Decker MD (2001) Combination vaccines. *Infect Dis Clin North Am* 15(1):209–230
14. Flanagan-Klygis EA, Sharp L, Frader JE (2005) Dismissing the family who refuses vaccines: A study of pediatrician attitudes. *Arch Pediatr Adolesc Med* 159(10):929–934
15. Garey MR, Johnson DS (1979) Computers and intractability: A guide to the theory of NP-completeness. W.H. Freeman, New York
16. Hall SN (2006) The design and analysis of pediatric vaccine formularies: theory and practice. University of Illinois, Urbana, Illinois Ph.D. Dissertation
17. Hochbaum DS (ed) (1997) Approximation algorithms for NP-hard problems. PWS, Boston, Massachusetts
18. Infectious Diseases in Children (2002) Almost 200 new drugs in development for use in children. *15(10):41–44*
19. Jacobson RM, Targonski PV, Poland GA (2007) Meta-analyses in vaccinology. *Vaccine* 25:3153–3159
20. Jacobson SH, Karnani T, Sewell EC (2003a) Analyzing the economic value of the hepatitis B-*Haemophilus influenzae* type B combination vaccine by reverse engineering a formulary selection algorithm. *Vaccine* 21:2169–2177
21. Jacobson SH, Karnani T, Sewell EC (2004) Assessing the impact of wastage on pediatric vaccine immunization formulary costs using a vaccine selection algorithm. *Vaccine* 22:2307–2315
22. Jacobson SH, Sewell EC (2002) Using Monte Carlo simulation to determine combination vaccine price distributions for childhood disease. *Health Care Manage Sci* 5:135–145
23. Jacobson SH, Sewell EC, Allwine DA, Medina EA, Weniger BG (2003b) Designing pediatric vaccine formularies and pricing pediatric combination vaccines using operations research models and algorithms. *Expert Rev Vaccines* 2(1):15–19
24. Jacobson SH, Sewell EC, Deuson R, Weniger BG (1999) An integer programming model for vaccine procurement and delivery for childhood immunization: a pilot study. *Health Care Manage Sci* 2:1–9
25. Jacobson SH, Sewell EC, Karnani T (2005) Engineering the economic value of two pediatric combination vaccines. *Health Care Manage Sci* 8(1):29–40
26. Jacobson SH, Sewell EC, Proano RA (2006) An analysis of the pediatric vaccine supply shortage problem. *Health Care Manage Sci* 9(4):371–389

27. Nemhauser GL, Wolsey LA (1999) Integer and combinatorial optimization. Wiley, New York
28. Srinivasan A (1999) Improved approximation guarantees for packing and covering integer programs. *SIAM J Comput* 29 (2):648–670
29. Sewell EC, Jacobson SH (2003) Using an integer programming model to determine the price of combination vaccines for childhood immunization. *Ann Oper Res* 119:261–284
30. Sewell EC, Jacobson SH, Weniger BG (2001) Reverse engineering a formulary selection algorithm to determine the economic value of pentavalent and hexavalent combination vaccines. *Pediatr Infect Dis J* 20(11):S45–S56
31. Weniger BG (1996) Economic analysis to meet the challenges of new combination vaccines. Presented at Vaccine Economics: Planning a Research Agenda for the Challenge of New and Improved Vaccines, Atlanta, GA
32. Weniger BG, Chen RT, Jacobson SH, Sewell EC, Deuson R, Livengood JR, Orenstein WA (1998) Addressing the challenges to immunization practice with an economic algorithm for vaccine selection. *Vaccine* 16(19):1885–1897