Management Science

Analyzing Screening Policies for Childhood Obesity

Yan Yang, Jeremy D. Goldhaber-Fiebert, Lawrence M. Wein,

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

Analyzing Screening Policies for Childhood Obesity

Yan Yang

Institute for Computational and Mathematical Engineering, Stanford University, Stanford, California 94305,
yanyang@stanford.edu

Jeremy D. Goldhaber-Fiebert

Centers for Health Policy and Primary Outcomes Research, Stanford University School of Medicine, Stanford, California 94305,
jeremygf@stanford.edu

Lawrence M. Wein

Graduate School of Business, Stanford University, Stanford, California 94305, lwein@stanford.edu

Because of the health and economic costs of childhood obesity, coupled with studies suggesting the benefits of comprehensive (dietary, physical activity, and behavioral counseling) intervention, the U.S. Preventive Services Task Force recently recommended childhood screening and intervention for obesity beginning at age 6. Using a longitudinal data set consisting of the body mass index of 3,164 children up to age 18 and another longitudinal data set containing the body mass index at ages 18 and 40 and the presence or absence of disease (hypertension and diabetes) at age 40 for 747 people, we formulate and numerically solve—separately for boys and girls—a dynamic programming problem for the optimal biennial (i.e., at ages 2, 4, . . . , 16) obesity screening thresholds. Unlike most screening problem formulations, we take a societal viewpoint, where the state of the system at each age is the population-wide probability density function of the body mass index. Compared to the biennial version of the task force's recommendation, the screening thresholds derived from the dynamic program achieve a relative reduction in disease prevalence of 3% at the same screening (and treatment) cost, or—because of the flatness of the disease versus screening trade-off curve—achieves the same disease prevalence at a 28% relative reduction in cost. Compared to the task force's policy, which uses the 95th percentile of body mass index (from cross-sectional growth charts tabulated by the Centers for Disease Control and Prevention) as the screening threshold for each age, the dynamic programming policy treats mostly 16-year-olds (including many who are not obese) and very few males under 14 years old. Although our results suggest that adult hypertension and diabetes are minimized by focusing childhood obesity screening and treatment on older adolescents, the shortcomings in the available data and the narrowness of the medical outcomes considered prevent us from making a recommendation about childhood obesity screening policies.

Key words: dynamic programming; applications; healthcare; statistics
History: Received September 26, 2011; accepted April 28, 2012, by Assaf Zeevi, stochastic models and simulation. Published online in *Articles in Advance* November 5, 2012.

1. Introduction

The prevalence of obesity in U.S. school-aged children and adolescents tripled during 1980–1999 and has plateaued since 1999 at ≈17%, with 31.7% of U.S. children ages 2–19 overweight or obese (Ogden et al. 2010). Obese children tend to have elevated lipid concentrations and blood pressure (Freedman et al. 2007) and often become obese adults (Singh et al. 2008) with an increased risk of diseases such as hypertension and type 2 diabetes (e.g., Hubert et al. 1983, Barrett-Connor 1989). This is not only a public health problem, but an economic (the cost of adult obesity has been estimated at $147 billion per year; Finkelstein et al. 2009) and national security (>25% of Americans ages 17–24 are unqualified for the military because of their weight; Mission: Readiness 2010) issue as well.

Two complementary approaches have been put forth to address this problem: a universal approach and a targeted approach. The universal approach focuses on better nutrition and more physical activity (as in First Lady Michelle Obama's "Let's Move" campaign; White House Task Force on Childood Obesity 2010) and involves schools, families, communities, and the food industry. Some school-based initiatives have generated reductions in obesity (Gortmaker et al. 1999, Coleman et al. 2005), though others have not shown statistically significant effects (HEALTHY Study Group 2010). A recent meta-analysis found that obesity reduction programs provide modest benefits, at least over the short run, including for children ages 0–5 (Waters et al. 2011).

The targeted approach requires measuring children's body mass index (BMI), which is the weight divided by the square of the height, measured in kg/m^2 and rounded to one decimal place. BMI is a practical and reasonably accurate (Reilly et al. 2000)

measure of a person's true body fat, and obesity and overweight are defined as being the 95th and 85th percentile of the gender- and age-based BMI distributions tabulated by the Centers for Disease Control and Prevention (CDC) (Kuczmarski et al. 2000). These tabulated percentages are based on historical distributions, which explains how 17% of U.S. children can exceed the 95th percentile. Many school districts and entire states require measuring a child's BMI every several years, but for surveillance—not screening—purposes (Nihiser et al. 2007).

Although an Institute of Medicine report states that healthcare professionals should routinely track children's BMI and offer evidence-based counseling and guidance (Institute of Medicine 2005), clinical trials for treating childhood obesity have been hampered until recently by small sample sizes, poor generalizability, and variable followup, leading to insufficient evidence for effectiveness of intervention in primary care settings (e.g., U.S. Preventive Services Task Force 2005). However, a targeted systematic review to support the updated U.S. Preventive Services Task Force (USPSTF) recommendations concludes that research into obesity treatment has improved in the past several years, leading to at least short-term benefits of comprehensive medium- to high-intensity (>25 hours over six months) interventions that include dietary, physical activity, and behavioral counseling components (Whitlock et al. 2010). Consequently, the USPSTF recommends childhood BMI screening and intervention for obesity starting at age six (USPSTF 2010).

This state of affairs highlights the need to assess possible intervention strategies. To our knowledge, the only study to do this compares screening and treatment at a single age (5, 10, or 15) to universal intervention, and suggests that, should effective universal interventions be available, this approach could lead to greater health benefits than the three single-age screening strategies, and that screening at age 15 is more beneficial than screening at age 5 or age 10 because some children who are obese at these younger ages are not obese at age 18 in the absence of treatment (Goldhaber-Fiebert et al. 2012).

In this paper, we derive optimal screening thresholds for children under 18 years old but do not prove that threshold policies are optimal among all policies. Using biennial BMI data from the National Longitudinal Survey of Youth (NLSY) samples (Bureau of Labor Statistics 2010), we model the Markovian evolution of the probability density function (pdf) of BMI for all U.S. children at ages $t = 0, 2, \ldots, 18$, assuming there is no obesity treatment. We optimize over the class of biennial threshold policies, where every child has his BMI measured every two years and receives treatment whenever his BMI exceeds a gender- and age-based threshold. The impact of treatment is estimated using the pooled analysis of three comprehensive moderate- or high-intensity behavioral interventions (Whitlock et al. 2010). The objective function of our model is to minimize the disease prevalence of these children when they reach age 40, which is estimated using Panel Study of Income Dynamics (PSID) data (Institute for Social Research 2010). The optimal screening problem is formulated as a deterministic dynamic program with an infinite-dimensional state (the pdf of the BMI in the population at age t). We then use a simulation model, which has much less restrictive assumptions than the dynamic program, to compare the performance of the USPSTF policy and the policy derived from the dynamic program. When we refer to an "optimal policy" we mean the solution to the dynamic program; because of shortcomings in the data and the problem formulation (which are discussed in §6), we are not suggesting that the dynamic programming policy is also the optimal approach to the actual problem. Our goal here is to present an initial mathematical formulation of this problem that makes the best use of available data and identifies key data needs and modeling assumptions that might improve future studies.

The problem considered here shares features with other screening and treatment problems, such as the screening of cholesterol (Garber et al. 1994), HIV viral load (D'Amato et al. 2000), CD4 count (Shechter et al. 2008), estimated glomerular filtration rate (Lee et al. 2008), or prostate specific antigen level (Lavieri et al. 2012). Whereas the latter four papers consider decisions of when to initiate or switch treatment from an individual's point of view (i.e., the state of the system measures a person's health status), our paper is perhaps closest in spirit to Garber et al. (1994) because both studies consider the time evolution of the population-wide pdf of a health measure (cholesterol in their case, BMI in our case). Whereas the goal of Garber et al. (1994) is forecasting future cholesterol levels in the absence of treatment, we embed the dynamics in the presence of screening and treatment into a dynamic programming framework and derive optimal screening thresholds.

Exploratory data analyses in §2 guide our problem formulation in §3. Parameter estimation is undertaken in §4 and a computational study appears in §5. The results are discussed in §6.

2. Exploratory Data Analyses

The NLSY data set has a longitudinal population cohort of children born to women who participated in the NLSY between 1970 and 1988. Each of the 3,164 children (1,566 boys and 1,598 girls) has biennial BMI measurements from ages 0 to 18, which are partially self-reported or mother reported. For children measured at odd-numbered ages, we linearly interpolate to obtain their BMI at even-numbered ages, and the analyses in this section considers BMI

at ages $0, 2, \ldots, 18$. Each child in this data set has an attached weight such that the weighted sample of the 3,164 children represents all children born between 1970 and 1988 to women ages 21–31 who were present in the United States in 1979, and all of our analyses are on the weighted sample.

We denote the BMI of a random child at age t by B_t and let B be the stochastic process $(B_0, B_2, \ldots, B_{18})$. We explore the Markov property of B in §2.1, the distribution of B_t in §2.2, and the distribution of the BMI increments $(B_{t+2} - B_t)$ in §2.3.

2.1. Markov Property of BMI

The mean and the standard deviation of B_t decrease during ages 0–4, plateau at ages 4–6, and increase during ages 6–18 (see Figure 1 of the electronic companion, available at http://faculty-gsb.stanford.edu/wein/personal/documents/obesity-ms-appendix.pdf), and hence we dismiss the possibility that B has stationary transition probabilities. We use the approach in Garber et al. (1994), which tests the hypothesis (all tests in this study are performed at the 95% confidence level) that B is a first-order Markov chain via a direct test of the Chapman–Kolmogorov equations. The significance levels of the chi-square statistics, which measure the difference between the observed and expected two-step (i.e., from t to $t+4$) transitions, are derived by a bootstrap technique (see §1 of the electronic companion for details). Of the 16 tests (ages $t = 0, 2, \ldots, 14$ for each gender), we reject the null hypothesis that BMI levels follow a first-order Markov chain in four cases: males at age 12 and females at ages 0, 6, and 14. Under a Bonferroni correction, which considers the $1 - 0.05/8 = 0.99625$ fractile of the chi-square distributions because there are eight age groups being tested, the Markov hypothesis is not rejected for any age–gender combination. However, because we are trying to retain the null hypothesis, this Bonferroni approach is not conservative.

2.2. Marginal Distribution of BMI

Initial exploratory analysis shows that B_t follows neither a normal nor a lognormal distribution for any t, although it is much closer to the latter (data not shown). Consequently, to increase the normality of the data, we employ a Box–Cox transformation of $\ln B_t$, which we denote by $X_t = ((\ln B_t)^\lambda - 1)/\lambda$ (Box and Cox 1964). The optimal value (i.e., the value that makes the distribution closest to normal) of the Box–Cox parameter is $\lambda = -2.15$ for males and $\lambda = -2.02$ for females. The resulting Q–Q plots in Figure 2 of the electronic companion suggest that X_t is reasonably approximated by a normal distribution, particularly for the higher age groups. The Kolmogorov–Smirnov goodness-of-fit test for a normal distribution shows that at the 95% confidence level, X_t passes four of the 20 tests: girls at ages 0, 8, 10, and 12. Further analysis shows that the majority of the deviation in the failed tests occurs in the left tail of the distribution (i.e., unusually low BMI levels), which is not important for the purposes of obesity screening. We also estimated the normal parameters from just the right half of X_t (i.e., only the data larger than the median), and obtained qualitatively similar results to those in Table 1 (data not shown).

2.3. BMI Increments

Because the increment $X_{t+2} - X_t$ depends on the value of X_t, we assess the normality of the increments using a 95% Kolmogrov–Smirnov test on the increment from age t to $t+2$ for $t = 0, 2, \ldots, 16$, and where the (untransformed) BMI values are segregated into different buckets. The results are given in Tables 6 and 7 in the electronic companion when BMI is divided into 25 buckets (<12, 12–13, 13–14, $\ldots$, 34–35, 35+). A cruder system with six buckets (<16, 16–18, 18–20, 20–22, 22–24, >24) generates the results in Tables 8 and 9 in the electronic companion. With 25 buckets, only 13 of the 362 tests (88 of the 450 possible tests had no data) failed the normality test, and 36 of the 108 normality tests failed in the six-bucket system.

Table 1 **For a Variety of Policies, the Treatment Prevalence p from (3) and the Hypertension and Diabetes Prevalences at Age 40**

Policy	Prevalences for males			Prevalences for females		
	Treatment	Hypertension	Diabetes	Treatment	Hypertension	Diabetes
No screening	0	0.2214	0.0636	0	0.2350	0.0568
Biennial USPSTF	0.1027	0.2076	0.0566	0.0850	0.2020	0.0467
Optimal: age 16	0.1027	0.2065	0.0566	0.0850	0.2003	0.0474
Optimal: ages (6, 16)	0.1027	0.2050	0.0559	0.0850	0.2003	0.0474
Optimal: ages (10, 16)	0.1027	0.2049	0.0558	0.0850	0.1987	0.0467
Optimal: ages (14, 16)	0.1027	0.2035	0.0548	0.0850	0.1941	0.0451
Optimal: ages (12, 14, 16)	0.1027	0.2035	0.0548	0.0850	0.1931	0.0447
Approximate DP: all ages	0.1027	0.2035	0.0548	0.0850	0.1930	0.0446
Optimal: ages (12, 14, 16)	0.2500	0.1869	0.0480	0.2500	0.1527	0.0358

Note. The optimal policies are derived from the dynamic program (DP) using the listed ages, the exhaustive search policy uses only the simulation model, and the approximate DP policy uses the algorithm in Bertsekas (2005).

3. Problem Formulation

By our analysis in §2.1, we assume that $X = (X_0, X_2, \ldots, X_{18})$ is a Markov process, where $X_t = ((\ln B_t)^\lambda - 1)/\lambda$ is the Box–Cox transformation of $\ln B_t$. Because we are considering this screening problem from a societal—rather than individual—point of view, the system state at age t is given by $f_t(x)$, which is the pdf of X_t. If treatment is given at age t, it is given after the measurement $X_t = x$. In the absence of treatment at age t, the increment $X_{t+2} - X_t$ given $X_t = x$ has pdf $g_{xt}(w)$, where w is the dummy variable for the value of the increment. By the analysis in §2.3, we assume that $g_{xt}(w)$ is the pdf of a normal distribution with mean μ_{gxt} and standard deviation σ_{gxt}. If treatment is given at age t, we assume that $X_{t+2} - X_t$ given $X_t = x$ has a normal distribution with pdf $h_{xt}(w)$, mean μ_{hxt}, and standard deviation σ_{hxt}; note that the impact of treatment is assumed to be independent of the impact of treatment received at previous ages. We consider biennial threshold policies with thresholds θ_t, where a person with $B_t > \theta_t$ (i.e., $X_t > ((\ln \theta_t)^\lambda - 1)/\lambda$) receives treatment at age t, for $t = 0, 2, \ldots, 16$ (although we define θ_0 for notational convenience, we do not allow treatment at birth and set $\theta_0 = \infty$). The system state equation is

$$f_{t+2}(x)$$
$$= \int_0^{((\ln \theta_t)^\lambda - 1)/\lambda} f_t(w) g_{wt}(x - w) \, dw$$
$$+ \int_{((\ln \theta_t)^\lambda - 1)/\lambda}^{\infty} f_t(w) h_{wt}(x - w) \, dw \quad \text{for } x \geq 0. \quad (1)$$

In the absence of any treatment (i.e., $\theta_t = \infty \; \forall t$), Equation (1) implies that if the distribution of X_0 is normal then the distribution of X_t is normal for $t = 2, 4, \ldots, 18$, which is consistent with our analysis in §§2.2 and 2.3 that concluded that both X_t and $X_{t+2} - X_t$ are roughly normal.

We assume that the initial system state, $f_0(x)$, is a skew-normal pdf with location parameter μ_{f_0}, scale parameter σ_{f_0}, and shape parameter α_{f_0}, which has pdf $f_0(x) = (2/\sigma_{f_0}) \phi((x - \mu_{f_0})/\sigma_{f_0}) \Phi((\alpha_{f_0}(x - \mu_{f_0}))/\sigma_{f_0})$, where $\phi(z)$ and $\Phi(z)$ are the pdf and cdf (cumulative distribution function) of the standard normal distribution (O'Hagan and Leonard 1976). The shape parameter α_{f_0} indicates the amount of skewness, and $f_0(x)$ reduces to a normal pdf when $\alpha_{f_0} = 0$.

For $t = 0, 2, \ldots, 18$, we define the disease function $c_t(x)$ to be the contribution to the disease prevalence at age 40 due to having $X_t = x$ at age t. The objective is to minimize the disease prevalence at age 40, which in §4.4 is computed using a nonseparable function of $(f_0(x), f_2(x), \ldots, f_{18}(x))$. We use a surrogate objective function,

$$\min_{\theta_0, \theta_2, \ldots, \theta_{16}} \sum_{t=0}^{9} \int_0^{\infty} c_{2t}(x) f_{2t}(x) \, dx, \quad (2)$$

which is separable, and hence fits into the dynamic programming framework. This issue is discussed further at the end of §4.4.

In §4.4, we estimate the disease functions $(c_0(x), c_2(x), \ldots, c_{18}(x))$ by assuming that the Markov property of X continues until age 40 (i.e., X_{40} depends on $(X_0, X_2, \ldots, X_{18})$ only via X_{18}) and by assuming that the prevalence of disease at age 40 is a function of the average BMI throughout the first 40 years; this latter assumption is relaxed in the sensitivity analyses in §5.3.

Because the USPSTF found no evidence of adverse effects from treatment on growth, eating-disorder pathology, or mental health (USPSTF 2010), the main costs associated with screening and treatment are financial. We simply impose a constraint on the fraction of children ages $2, 4, \ldots, 16$ that screen positive (i.e., have a BMI exceeding their gender- and age-specific BMI threshold) each year, assuming there are the same number of children of each age in the population:

$$\frac{1}{8} \sum_{t=1}^{8} \int_{((\ln \theta_{2t})^\lambda - 1)/\lambda}^{\infty} f_{2t}(x) \, dx \leq p. \quad (3)$$

We refer to p as the *treatment prevalence*. Because the screening cost is negligible compared to the treatment cost, the left side of (3) is essentially proportional to the annual treatment cost. By varying p on the right side of (3), we can generate a range of policies and compare derived policies to recommended policies (e.g., USPSTF 2010).

Notice that although our system state is a pdf, the state transition is deterministic, and hence our problem, which is given by (1)–(3), is a deterministic constrained dynamic program. We do not derive any structural results about this dynamic program; e.g., we do not show that the optimal policy is a threshold policy.

4. Parameter Estimation

The initial BMI distribution, the BMI increments in the absence of treatment, the BMI increments in the presence of treatment, and the disease functions are estimated in §§4.1–4.4.

4.1. Initial BMI

Using the BMI data at birth, the maximum-likelihood estimation parameter values for the skew-normal pdf $f_0(x)$ are $\mu_{f_0} = 0.42$, $\sigma_{f_0} = 0.009$, and $\alpha_{f_0} = -1.00$ for boys, and $\mu_{f_0} = 0.43$, $\sigma_{f_0} = 0.006$, and $\alpha_{f_0} = -1.30$ for girls. As expected from §2.2, these two distributions have very little skewness (figures not shown).

4.2. BMI Increments

In the absence of treatment, we assume that $X_{t+2} - X_t$ given $X_t = x$ is normal with pdf $g_{xt}(w)$. We denote its mean and standard deviation by μ_{gxt} and σ_{gxt}, and

we estimate the values of these parameters using the NLSY data set. These parameter values are plotted as functions of x for $t = 0, 2, \ldots, 16$ for boys and girls in Figure 3 of the electronic companion.

4.3. Treatment

If treatment is given at age t, we assume that $X_{t+2} - X_t$ given $X_t = x$ is normal with pdf $h_{xt}(w)$, mean μ_{hxt}, and standard deviation σ_{hxt}, where this increment captures the natural progression in the absence of treatment plus the effect of treatment. Although there are a number of meta-analyses that assess the impact of treatment for childhood obesity, we use data from the three comprehensive moderate- or high-intensity programs considered in Table 1 of Whitlock et al. (2010) (Savoye et al. 2007, Reinehr et al. 2006, Nemet et al. 2005), which is a targeted systematic review to support the updated USPSTF recommendations. These data provide aggregate results on the effect of treatment and do not allow us to reliably estimate the dependence of treatment on the gender, age, or pretreatment BMI level; the latter two factors are interrelated because younger children tend to have smaller BMI values (e.g., the CDC obesity threshold at 16 years old is approximately 50% larger than at 6 years old).

The approach used to model the effect of treatment in the face of limited data is both important (i.e., it will affect the qualitative conclusions of our study) and nonobvious. Given that we cannot develop detailed functions that depend on age or pretreatment BMI level, three modeling approaches come to mind: let the treatment-induced change in (i) BMI, B_t; (ii) transformed BMI, X_t; or (iii) z-BMI be independent of gender, age, and pretreatment BMI level. We denote z-BMI by $Z_t = ((B_t/M_t)^{L_t} - 1)/(L_t S_t)$, where M_t is the median, S_t is the generalized coefficient of variation, and L_t is the parameter of the Box–Cox transformation (these three parameters also depend on gender and are given in Kuczmarski et al. 2000); Z_t can be converted into percentiles using standard normal distribution tables. We pursue option (ii), and in §3 of the electronic companion, we show that our treatment model accurately captures the dependence on pretreatment BMI (albeit in a somewhat narrow range of pretreatment levels) and successfully predicts that the change in z-BMI is larger for younger children than for older children (as shown in clinical studies by Epstein et al. 2007).

We estimate the impact of treatment in three steps (details are in §2 of the electronic companion). First, we transform the BMI data in Table 1 of Whitlock et al. (2010) from B to X using a Taylor series expansion around the mean of B, which gives us the mean and standard deviation of the change in X due to treatment for each of the three studies. Second, we compute the mean and standard deviation of the net treatment (i.e., treatment minus control) effect, pooled

over the three studies. Third, we combine the natural progression and the net treatment effect to obtain μ_{hxt} and σ_{hxt}.

4.4. Disease Functions

The most difficult task in this section is to estimate the four sets of disease functions ($c_0(x_0)$, $c_2(x_2)$, $\ldots, c_{18}(x_{18})$), which allow us to predict the disease (hypertension or diabetes) prevalence at age 40 in terms of ($B_0, B_2, \ldots, B_{18}$) for each gender. We use three different data sets, which measure BMI from three different sets of individuals to estimate these functions: the NLSY data that contain ($B_0, \ldots, B_{18}$); the PSID data (Institute for Social Research 2010), which have self-reported BMI at ages 18 and 40 and the presence or absence of hypertension and of type 2 diabetes by age 40; and the National Health and Nutrition Examination Survey (NHANES) data (National Center for Health Statistics 2006), which contain annual BMI data for ages [18, 40]. The NHANES data set is cross-sectional, and to use it to infer patterns of BMI changes between ages 18 and 40 in the PSID data requires the assumption that there are no substantial cohort-specific trends.

This subsection begins with a literature review, which motivates our key assumption that determines the independent variable in our logistic regression for the prevalence of disease at age 40. Then we use the NHANES database to estimate the parameter γ, which tells us how to weight B_{18} and B_{40} in the PSID data so as to succinctly capture how BMI changes between ages 18 and 40. With γ in hand, we reexpress the independent variable in terms of X_t rather than B_t to obtain the disease functions and run four logistic regressions (for each combination of gender and disease) by stringing together sample paths from the NLSY and PSID data sets. Finally, we use the regression results to calculate the disease prevalence at age 40 for a generic set of BMI thresholds.

4.4.1. Literature Review. We briefly review the literature on how BMI throughout various stages of life impacts adult disease. A key issue—whether, after controlling for adult BMI, childhood BMI impacts adult health—has not been answered conclusively. A 55-year follow-up study found that, after controlling for adult BMI and other factors such as smoking and socioeconomic status, the risk of heart disease mortality among men increased with adolescent BMI, with many of the associated deaths occurring in their 40s (Must et al. 1992). Although another study (Mahoney et al. 1996) obtained similar results, four studies found no influence on adult disease from childhood BMI after controlling for adult BMI. However, these four studies do not necessarily conflict with the finding in Must et al. (1992): Colditz et al. (1990) considered diabetes in

women, Willett et al. (1995) considered heart disease in women, Freedman et al. (2001) followed up at ages 18–37, and Raitakari et al. (2003) followed up at ages 24–39. On a related issue, some studies have found that weight gain or BMI increase in adulthood is a strong predictor of adult disease (Wilsgaard and Arnesen 2007 and references therein).

Several studies have shown that adult disease is a *j*-shaped function of childhood weight, where low-weight children were undernourished, leading to adult obesity (Abraham et al. 1971, Gunnell et al. 1998). Another study showed that death by age 32 is a *j*-shaped function of B_{18}, with lung cancer and cardiovascular disease being the main cause of death for low- and high-BMI, respectively.

Obesity at a fixed age in childhood and young adulthood is associated with risk factors for heart disease at that age (e.g., Freedman et al. 2007). Similarly, childhood obesity is a strong predictor of young-onset type 2 diabetes (Franks et al. 2007). Moreover, childhood obesity is associated with fibrous plaques (Berenson et al. 1998) and coronary artery calcification (Mahoney et al. 1996) in young adults, suggesting that "the development of coronary heart disease may depend on the cumulative lifetime effects of obesity" (Freedman et al. 2001, p. 715). On a different note, low BMI up to two years of age followed by an early adiposity rebound is associated with diabetes in young adulthood (Eriksson et al. 2003, Bhargava et al. 2004); we do not model this effect, but we discuss it further in §6.

4.4.2. Key Assumption. Because of the conceptual simplicity of the quote from Freedman et al. (2001) and for lack of definitive data, we estimate the disease functions under the assumption that *the prevalence of disease at age* 40 *depends on the average* (*or, equivalently, cumulative*) *amount of obesity throughout the first* 40 *years of life*. We emphasize that we cannot test this assumption directly, but rather we assume that it is true for the purposes of analysis; as discussed in §6, the lack of an evidence base for this assumption—or any competing assumptions—prevents us from making policy recommendations. To operationalize this assumption, we take as our measure of obesity at age t the BMI B_t. Before carrying out the remainder of these calculations, we note that the *z*-BMI Z_t may be a more appropriate measure of obesity than B_t because of the increase with age of the population-wide mean BMI. Consequently, we consider a variant of the key assumption where our measure of obesity is Z_t instead of B_t in §5.3, where we find that B_t is a more powerful predictor of adult disease incidence (as measured by McFadden pseudo-R^2; McFadden 1973) than Z_t.

Recall that the NLSY data has B_t for $t = 0, 2, \ldots, 18$, and the PSID data set has B_{18}, B_{40}, and D_{40} (which equals 1 or 0 in the presence or absence of disease) for a different set of individuals. Although it

would be natural to use $\sum_{t=0}^{20} B_{2t}$ as a measure of the cumulative amount of obesity up to age 40, even if we string together the sample paths from the NLSY and PSID data sets, we still only have BMI data for $(B_0, B_2, \ldots, B_{18}, B_{40})$. Hence, we use the NHANES data to approximate the BMI evolution over ages $[18, 40]$ by a convex combination of B_{18} and B_{40}. If we define $p_{40} = P(D_{40} = 1)$, then—based on our key assumption—we can run a logistic regression where $\ln(p_{40}/(1 - p_{40}))$ is the dependent variable and the average BMI throughout ages $[0, 40]$ is the independent variable. This latter quantity is given by

$$\frac{18}{40} \frac{\sum_{t=0}^8 B_{2t} + B_{18,N}}{10} + \frac{22}{40} [\gamma B_{18,P} + (1 - \gamma) B_{40}], \quad (4)$$

where the subscripts N and P are used to denote the NLSY and PSID data sets, respectively, because we have B_{18} values from both data sets (note that in both the dependent and independent variables, we are omitting a subscript for each individual person).

4.4.3. Estimating γ. To find γ in (4), we use the NHANES data set (National Center for Health Statistics 2006), which contains annual BMI data for ages $[18, 40]$ for 3,721 males and 4,245 females. The mean BMI for each age is plotted in Figure 5 in the electronic companion, along with the best-fitting quadratic curve, which has endpoints $\hat{\mu}_{f,18}$ and $\hat{\mu}_{f,40}$. We compute γ by equating an estimate of the average BMI during $[18, 40]$, which is the area under the quadratic curve divided by 22, to the weighted average $[\gamma \hat{\mu}_{f,18} + (1 - \gamma) \hat{\mu}_{f,40}]$. This procedure yields $\gamma = 0.49$ for females and $\gamma = 0.23$ for males. Note that $\gamma \approx 0$ if the additional weight during $[18, 40]$ is gained soon after age 18, and $\gamma \approx 1$ if the additional weight is gained just before age 40. Hence, our estimates suggest that females tend to gain weight uniformly throughout this age range (we do not exclude pregnant women from this analysis), whereas males tend to increase their BMI in the earlier years of this age range.

4.4.4. Conversion from B_t to X_t. Note that for the purposes of solving the dynamic program, it suffices to minimize the independent variable (i.e., the independent variable in (4) can be taken as the cost function); the regression coefficients are needed to estimate D_{40} under the optimal policy but are not needed to derive the optimal screening and treatment policy. Now that we have the independent variable in terms of B_t, we need to convert it into functions of $X_t = ((\ln B_t)^\lambda - 1)/\lambda$ for $t = 0, 2, \ldots, 18$. By (4), the cost function can be given by

$$\frac{18}{40} \frac{\sum_{t=0}^8 \exp[(\lambda X_{2t} + 1)^{1/\lambda}] + \exp[(\lambda X_{18,N} + 1)^{1/\lambda}]}{10}$$

$$+ \frac{22}{40} [\gamma \exp[(\lambda X_{18,P} + 1)^{1/\lambda}]$$

$$+ (1 - \gamma) \exp[(\lambda X_{40} + 1)^{1/\lambda}]]. \quad (5)$$

To express (5) solely in terms of $(X_0, X_2, \ldots, X_{18})$, we use the PSID data set to write B_{40} as a quadratic function of B_{18}, which yields $B_{40} = a_2 B_{18,N}^2 + a_1 B_{18,N} + a_0$, where $(a_2, a_1, a_0) = (-0.0198, 1.94, -7.27)$ for males and $(-0.0210, 2.19, -11.19)$ for females.

By (5) and by factoring out 1/40 and dropping the data set subscript from X_{18}, the disease functions are

$$c_t(x_t) = \tfrac{9}{5} \exp[(\lambda X_t + 1)^{1/\lambda}] \quad \text{for } t = 0, 2, \ldots, 16, \quad (6)$$

$$c_{18}(x_{18}) = \left(\tfrac{9}{5} + 22\gamma\right) \exp[(\lambda X_{18} + 1)^{1/\lambda}]$$
$$+ 22(1 - \gamma)(a_2 \exp[2(\lambda X_{18} + 1)^{1/\lambda}]$$
$$+ a_1 \exp[(\lambda X_{18} + 1)^{1/\lambda}] + a_0), \quad (7)$$

where $(\lambda, \gamma, a_2, a_1, a_0) = (-2.15, 0.23, -0.0198, 1.94, -7.27)$ for males and $(-2.02, 0.49, -0.0210, 2.19, -11.19)$ for females.

4.4.5. Logistic Regression. We now run the logistic regression, where the dependent variable is $\ln(p_{40}/(1 - p_{40}))$ and the independent variable is (4). To review, the BMI values in the independent variable are $(B_0, B_2, \ldots, B_{16}, B_{18,N}, B_{18,P}, B_{40})$, where we have $(B_0, B_2, \ldots, B_{16}, B_{18,N})$ for 1,566 boys and 1,598 girls from the NLSY data set, and we have $(B_{18,P}, B_{40})$ for 279 men and 468 women from the PSID data set. Using the Markov assumption that B_{40} depends on $(B_0, B_2, \ldots, B_{18})$ only via B_{18}, we string the data sets together by randomly assigning sample paths $(B_0, B_2, \ldots, B_{18,N})$ from the NLSY data based on the value of $B_{18,P}$ in the PSID data, although we employ bootstrapping to overcome the limited sample size of the latter data set. More specifically, the PSID data for 747 people are sampled with replacement with sampling probability proportional to the sample weights (which describe how representative each of the 747 people are in the general population), and create a bootstrapped data set of 747,000 people. For each bootstrap sample $(B_{18,P}, B_{40}, D_{40})$, one NLSY sample path $(B_0, B_2, \ldots, B_{16}, B_{18,N})$ of the same gender is selected to link to it with a probability that is proportional to $\phi((B_{18,P} - B_{18,N})/0.03)$, where $\phi(x)$ is the standard normal pdf and the standard deviation of 0.03 induces a strong distance effect while maintaining a reasonable number of NLSY samples in each selection. The results of the four logistic regressions are given in Table 12 of the electronic companion.

4.4.6. Computing the Disease Prevalence. Using Equation (1), we simulate 10^6 sample paths of BMI values $(x_0, x_2, \ldots, x_{18})$ under a generic threshold policy $(\theta_0, \theta_2, \ldots, \theta_{16})$. By construction, note that the independent variable in (4) can also be expressed as $(1/40) \sum_{t=0}^{9} c_{2t}(x_{2t})$. For each individual, we calculate his value of the independent variable using (6) and (7), and use the regression coefficients in Table 12

of the electronic companion to determine his value of p_{40}, which is

$$\frac{\exp(\beta_0 + (\beta_1/40) \sum_{t=0}^{9} c_{2t}(x_{2t}))}{1 + \exp(\beta_0 + (\beta_1/40) \sum_{t=0}^{9} c_{2t}(x_{2t}))}. \quad (8)$$

We compute the disease prevalence by taking the average of p_{40} over all 10^6 individuals (where each individual has a BMI sample path). From a computational viewpoint, this approach is much easier than integrating (8) over the 10-dimensional joint pdf of $f_t(x)$.

4.4.7. The Surrogate Objective Function. Note that our disease prevalence is the expected value of the quantity in (8), which is a nonseparable function of $(f_0(x), f_2(x), \ldots, f_{18}(x))$, and hence cannot be used in the dynamic programming formulation. Instead, we minimize the separable cost function in Equation (2). Although we have not been able to show analytically that minimizing (2) leads to minimizing the simulated disease prevalence in (8), the simulated disease prevalence is indeed monotonic in $\sum_{t=0}^{9} \int_0^\infty c_{2t}(x) f_{2t}(x) \, dx$ in our numerical computations in §5.2.

5. Computational Results
We briefly discuss the computational aspects of the dynamic program in §5.1. The base-case results and the sensitivity analyses are presented in §§5.2 and 5.3, respectively.

5.1. The Dynamic Program
To deal with constraint (3), we use a "finite-fuel" approach (e.g., Benes et al. 1980), where $y_t = p - (\sum_{s=1}^{t/2-1} \int_{(\ln \theta_{2s})^\lambda - 1)/\lambda}^\infty f_{2s}(x) \, dx)/8$ is the slack in the treatment prevalence from age t to age 16, and augment the system state to be $(f_t(x), y_t)$. If we define $V_t(f_t(x), y_t)$ to be the minimum remaining cost until age 18 when the state at age t is $(f_t(x), y_t)$, then the optimality equation for $t = 2, 4, \ldots, 16$ is

$$V_t(f_t(x), y_t) = \min_{\theta_t} \left\{ \int_0^\infty c_t(x) f_t(x) \, dx \right.$$
$$\left. + V_{t+2}\left(f_{t+2}(x), y_t - \tfrac{1}{8} \int_{((\ln \theta_t)^\lambda - 1)/\lambda}^\infty f_t(x) \, dx\right) \right\}, \quad (9)$$

and

$$V_{18}(f_{18}(x), y_{18}) = \int_0^\infty c_{18}(x) f_{18}(x) \, dx, \quad (10)$$

where the minimization is subject to the finite-fuel constraint $y_{18} \geq 0$, and the state transitions for $f_t(x)$ are defined in (1). Note that the optimal thresholds are independent of disease (i.e., hypertension versus diabetes) because the disease prevalence for both diseases is an increasing function of the objective function in Equation (2).

Although our dynamic program is deterministic, its continuous state (i.e., $f_t(x)$) makes for unwieldy calculations, and we have been unable to solve for the optimal thresholds for more than three ages (i.e., we can only solve a three time-period dynamic program). Consequently, we also use an approximate dynamic programming algorithm, more specifically an extension of the rollout algorithm that is specifically tailored to our setting of discrete-time constrained deterministic dynamic programming problems (Bertsekas 2005). The algorithm, specified in §2 of Bertsekas (2005), relies on a suboptimal policy called the base heuristic, upon which it improves, and we use the optimal $(14, 16)$ policy (i.e., the optimal policy among those that screen only at ages 14 and 16) as the base heuristic.

5.2. Base-Case Results

In this subsection, we solve the dynamic program for a wide variety of treatment prevalences and compare these policies to a variant of the USPSTF recommendation. Recall that the USPSTF recommends that children be screened and treated for obesity starting at age 6 (USPSTF 2010). Although the USPSTF does not explicitly recommend a screening frequency, they suggest that screening be done during routine pediatric assessments, which typically occur every year or two. We adapt their recommendation to biennial screening, and refer to the biennial USPSTF policy as the one that sets the BMI thresholds θ_t equal to the 95th percentile of the BMI distributions in the CDC growth charts (Kuczmarski et al. 2000) for ages $t = 6, 8, \ldots, 16$ (and sets $\theta_t = \infty$ for $t = 0, 2, 4$). It is important to remember that the CDC BMI distributions do not reflect the BMI distributions in the NLSY data set: obesity has tripled in recent decades and approximately 15% of children in the NLSY data set exceed the 95th percentile of the CDC BMI distributions.

Because we have been able to solve the dynamic program for up to only three time periods (this takes several days on a personal computer, and incorporating each additional time period increases the running time by approximately a factor of 800, which is the size of our discretized state space), we present results that progressively restrict the number of ages that are tested to one, two, and three, respectively. The main results are presented as trade-off curves in Figure 1 as well as in Figures 6 and 7 in the electronic companion, respectively, with selected numerical results appearing in Table 1.

Figure 6 of the electronic companion has eight trade-off curves that test at a single age: $2, 4, \ldots, 16$. For each gender and disease, performance improves with increasing age (except for males at age 8). Figure 7 of the electronic companion displays trade-off curves for three two-age screening polices, at ages

$(6, 16)$, $(10, 16)$, and $(14, 16)$. Screening at ages $(14, 16)$ is the best among these three policies. Figure 1 reveals that the three-age policy using ages $(12, 14, 16)$ performs only marginally better than the optimal two-age policy.

The optimal $(12, 14, 16)$ policy in Table 1 has two potential sources of suboptimality: it uses the sum of BMIs as its separable objective function, rather than the disease prevalence computed by simulation via Equation (8), and the dynamic programming solution only screens at three ages. We perform two other calculations to assess these suboptimalities. First, an exhaustive search of the $(14, 16)$ policy using the simulation model yields exactly the same solution as the optimal $(14, 16)$ policy in Table 1 (however, an exhaustive search for the optimal three-stage policy is extremely onerous from a computational standpoint). Second, although the approximate dynamic programming procedure based on Bertsekas (2005) generates a policy that treats at more than three ages, it generates nearly the same disease prevalences as the optimal $(12, 14, 16)$ policy (Table 1). Taken together, all of these results strongly suggest that the optimal $(12, 14, 16)$ policy is nearly optimal using the simulation in Equation (8) among the class of all threshold policies (i.e., allowing screening at ages $2, 4, \ldots, 16$) for our set of parameter values.

Overall, the optimal trade-off curves are only slightly convex and are rather flat. The flatness of the trade-off curves leads to different qualitative conclusions regarding the magnitude of the performance gap between the optimal three-age policy and the biennial USPSTF policy (Figure 1 and Table 1): the optimal three-age policy provides a 3% mean relative reduction in disease prevalence (2.0% for male hypertension, 3.2% for male diabetes, 4.4% for female hypertension, and 4.3% for female diabetes) at the same treatment prevalence, or a 28% mean relative reduction in treatment prevalence (28.5% for male hypertension, 29.1% for male diabetes, 28.0% for female hypertension, and 27.3% for female diabetes) at the same disease prevalence.

Relative to no obesity screening, the optimal three-age policy with a treatment prevalence equal to that of the biennial USPSTF policy achieves a relative reduction in disease prevalence of 8.1% for male hypertension, 13.8% for male diabetes, 17.8% for female hypertension, and 21.3% for female diabetes. To calibrate the treatment prevalence, we note that if the thresholds were set at the overweight level (i.e., 85th percentile of the CDC distributions) for ages $6, 8, \ldots, 16$, then the treatment prevalence is 0.2158 for boys and 0.1930 for girls. As an upper bound, the optimal three-age policy with a treatment prevalence equal to 0.25 reduces disease prevalence (compared to no screening) by

Figure 1   Trade-Off Curves of Disease Prevalence at Age 40 vs. Treatment Prevalence (Equation (3)) for the Optimal Three-Age Screening Policy for Ages 12, 14, and 16 and the Optimal Two-Age Screening Policy for Ages 14 and 16 from Figure 7 of the Electronic Companion
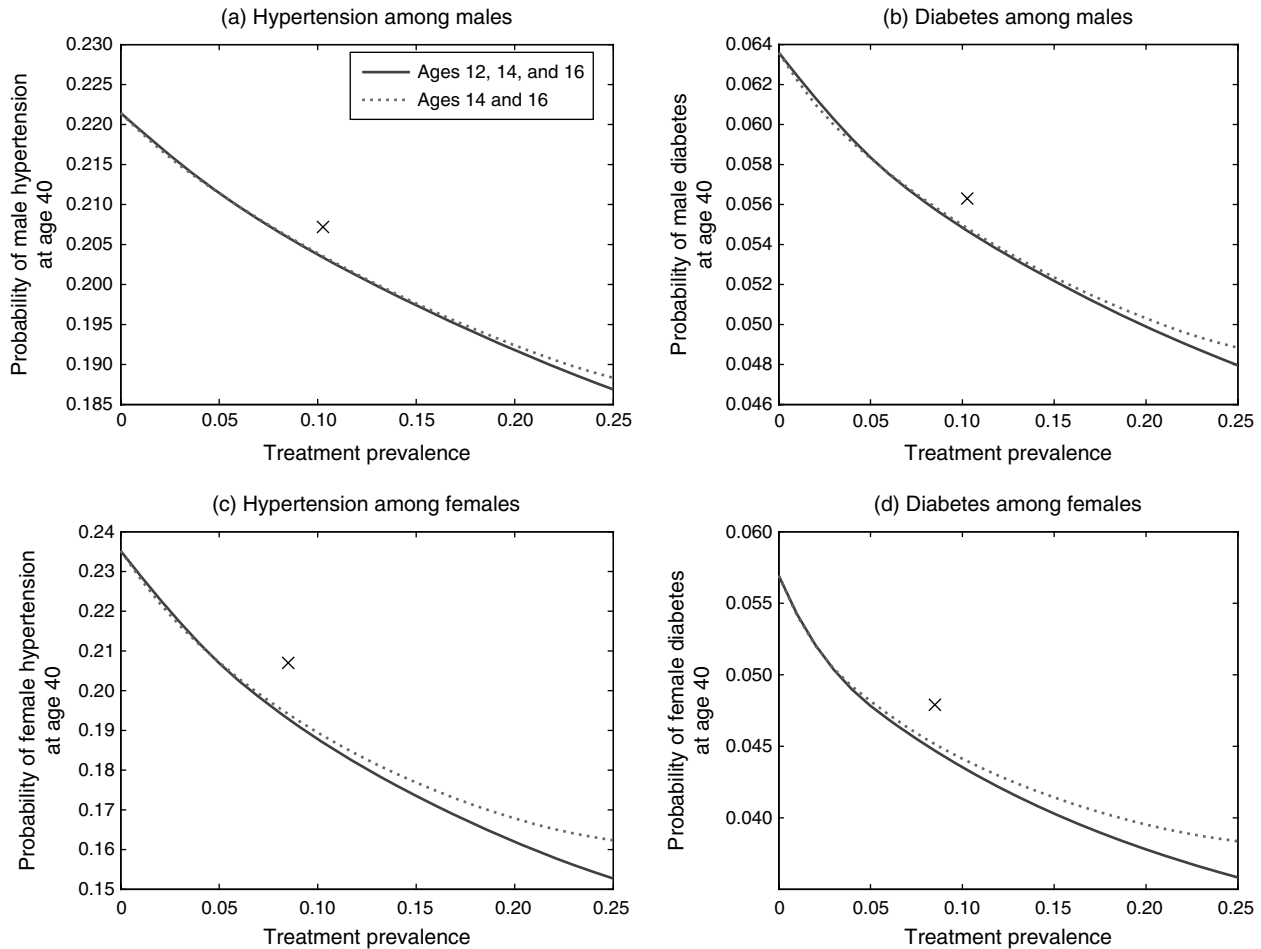
Note. The "×" symbol denotes the biennial version of the USPSTF recommendation.

15.6% for male hypertension, 24.5% for male diabetes, 35.0% for female hypertension, and 37.0% for female diabetes.

The optimal multiage policies achieve their improvements over the biennial USPSTF policy by being less selective—as measured by the threshold's corresponding percentile in the CDC BMI distributions—with increasing age; the same is not always true for the actual thresholds, e.g., the optimal BMI threshold is smaller at age 14 than age 16 in the optimal (14, 16) female policy. The majority ($\approx$51% for females and $\approx$70% for males) of treatment occurs at 16 years old (including both males and females who fall under the 85th CDC percentile, which is the overweight threshold), and very few ($\approx$1.7%) males under 14 years old are treated (Figure 2); these percentages are calculated using the system state pdf's $f_t(x)$ under the optimal solution, the values in Figure 2, and the CDC growth curves (i.e., we use these three pieces of information to compute the fraction of children of each age and gender that receive treatment).

5.3. Sensitivity Analyses

In this subsection, we carry out four sensitivity analyses: we allow treatment efficacy to be dependent on age and pretreatment BMI, and we consider two variants of the key assumption: disease prevalence depends on average z-BMI rather than on average BMI, and disease prevalence at age 40 depends on childhood BMI only via B_{18}.

5.3.1. Age-Dependent Treatment Efficacy. To assess the robustness of our results with respect to age-dependent treatment efficacy, we assume that treatment is more effective for children ages 6, 8, and 10 than for children ages 12, 14, and 16, and we consider the ratio of the change in untransformed BMI due to treatment for children ages 6, 8, and 10 divided by the change in untransformed BMI due to treatment for children ages 12, 14, and 16. This ratio is computed by replacing the mean net treatment effect, μ_T (-0.00183 for males and -0.00214 for females in §4.3), by $\mu_T - d$ for ages 6, 8, and 10, and $\mu_T + d$ for ages 12, 14, and 16, and choosing d to achieve

Figure 2   Optimal Screening Thresholds for the Three-Age Screening Policy at Ages 12, 14, and 16
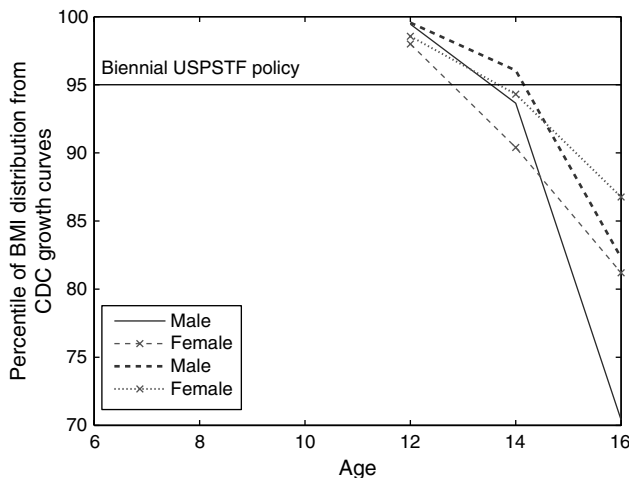
Notes. Thresholds are expressed as the percentile of the gender- and age-based BMI distributions tabulated by the CDC (Kuczmarski et al. 2000), for two cases: The treatment prevalence equals that of the biennial USPSTF policy (—— for males and -×- for females) and the disease prevalence equals that of the biennial USPSTF policy (··· for males and ··×·· for females). The biennial USPSTF policy uses the 95th percentile for ages 6, 8, . . . , 16.

the desired ratio, after converting from X_t to B_t and averaging over the top 15% of NLSY BMI values for each age, which is the subset of children who are apt to receive treatment. Under this construction, the mean and variance of the treatment effect for all children is independent of d. Note that when $d = 0$, which corresponds to our base case, this ratio is 0.61 for females and 0.62 for males (see Figure 4 of the electronic companion).

We now consider two policies, the optimal two-stage $(10, 16)$ policy and the optimal two-stage $(14, 16)$ policy, and compute the breakeven value of the ratio in the previous paragraph that equates the performance of these two policies. The breakeven ratio is 0.92 for males and 0.85 for females. If instead we compare the biennial USPSTF policy and the optimal $(14, 16)$ policy, the breakeven values are 1.03 for males and 0.94 for females, which are approximately 60% higher than the base-case values (i.e., the young-to-old ratio of the reduction in BMI needs to be 60% higher in the base case for these two policies to achieve the same performance).

5.3.2. BMI-Dependent Treatment Efficacy. We perform a similar sensitivity analysis as in §5.3.1, but assume that bigger treatment effects occur for children with higher pretreatment BMI levels rather than for younger children. We let the net treatment effect be $\mu_T - 0.441d$ for males ($\mu_T - 0.856d$ for females) with pretreatment BMI levels above the 97th CDC percentile, and be $\mu_T + d$ for those below the 97th percentile (0.441 and 0.856 maintain the mean treatment effect at 0 for the top 15% of NLSY BMI values), where

there is a one-to-one correspondence between d and the ratio of the change in untransformed BMI due to treatment for children above the 97th percentile divided by the change in untransformed BMI due to treatment for children between the 85th NLSY percentile and the 97th CDC percentile. When $d = 0$, this ratio is 1.54 for males and females (see Figure 4 of the electronic companion). We find that the breakeven ratio that equates the performance of the biennial USPSTF policy and the optimal $(14, 16)$ policy is 5.86 for males and 5.14 for females, which are much larger than the corresponding ratios in the sensitivity analysis for age-dependent treatment efficacy.

5.3.3. Obesity Measured as z-BMI Rather than BMI. Here, we reestimate the disease functions in §4.4 but assume that obesity at age t is measured by the z-BMI Z_t rather than B_t. We set $\gamma = 0.5$ (which generates higher pseudo-R^2 values than the γ values used earlier) and replace B_t by Z_t in Equation (4). Because the CDC charts are not tabulated through age 40, to compute the values of M_{40}, S_{40}, and L_{40}, we estimate these parameters from the PSID age 40 data, and then modify them by the ratios of the age 18 parameters derived from the CDC charts divided by the age 18 parameters derived from the PSID data. The new McFadden pseudo-R^2 values are $(0.020, 0.032, 0.100, 0.080)$, which are smaller than the values in Table 11 of the electronic companion. Overall, the results are qualitatively similar to the base-case results, in that the optimal $(12, 14, 16)$ policy provides a 3% reduction in disease prevalence relative to the biennial USPSTF policy. The main difference is that the optimal single-age policy (i.e., test only at age 16) is nearly optimal compared to the policy that screens at ages 12, 14, and 16.

5.3.4. Adult Disease Independent of BMI Before Age 18. Finally, we forego the key assumption that disease prevalence at age 40 depends on the cumulative lifetime obesity, and we consider the extreme case in which the BMI at age 18 is the sole childhood factor (aside from gender) that affects the disease prevalence at age 40. More specifically, the objective function of our dynamic program is changed from (2) to minimizing $\int_0^\infty x f_{18}(x)\, dx$. Not surprisingly, this change causes a shift toward more treatment at age 16 and less treatment at earlier ages. For example, in Table 1 the CDC threshold percentiles for the optimal $(14, 16)$ policy for males are 90.8 for age 14 and 74.1 for age 16, and they change to 96.1 for age 14 and 62.1 for age 16 under the new objective function. For females, the $(14, 16)$ percentiles change from $(85.7, 81.6)$ to $(93.9, 72.8)$.

6. Discussion

Our primary result is that—relative to a biennial version of the USPSTF policy–the optimal policy is predicted to achieve a 3% reduction in hypertension or

diabetes prevalence at age 40 at the same cost or achieve the same disease prevalence at a 28% reduction in cost. Whereas the USPSTF biennial policy sets thresholds at the 95th percentile of the CDC BMI distributions for ages $6, 8, \ldots, 16$, the optimal policy treats mostly 16-year-olds and few children under age 14. This is largely because of the high false positive rate associated with treating younger children (Goldhaber-Fiebert et al. 2012). More specifically, additional calculations (averaged over both genders) with our simulation model show that, in the absence of treatment, 26.8% of obese 6-year-olds, 18.8% of nonobese 6-year-olds, 39.4% of obese 16-year-olds, and 16.9% of nonobese 16-year-olds develop hypertension by age 40. The corresponding percentages for diabetes at age 40 are 7.6%, 4.8%, 12.2%, and 4.2%, respectively. If obese 6-year-olds are treated at age 6, 25.1% have hypertension and 6.9% have diabetes at age 40, and if obese 16-year-olds are treated at age 16, 34.9% have hypertension and 10.2% have diabetes at age 40. These calculations allow for three observations: (1) obese 16-year-olds are more likely than obese 6-year-olds to have disease at age 40 in the absence of treatment, (2) treatment of an obese 16-year-old at age 16 reduces the disease prevalence at age 40 more than the treatment of an obese 6-year-old at age 6, and (3) nonobese 6-year-olds are more likely to have disease at age 40 than nonobese 16-year-olds in the absence of treatment.

In our view, the total amount of screening to perform (i.e., setting the treatment prevalence p in Equation (3)) is not clear-cut because the trade-off curve is only modestly convex: the absolute value of the slope of the optimal trade-off curve in Figure 1, which is a measure of cost-effectiveness, decreases by an average of 72% as the treatment prevalence increases from 0 to 0.25. Even with perfect compliance, the trade-off curve in Figure 1 is somewhat flat: it takes roughly 20 childhood treatments (each taking >25 hours over six months) to avert a single case of disease by age 40 (referring to Table 1 and conservatively assuming that no adults get both hypertension and diabetes, the number of treatments per averted case is $[(0.2214 - 0.2035) + (0.0636 - 0.0548)]/(8 \times 0.1027) = 30.8$ for boys and $[(0.2350 - 0.1930) + (0.0568 - 0446)]/(8 \times 0.085) = 12.5$ for girls). At the treatment prevalence of the biennial USPSTF policy, the optimal $(12, 14, 16)$ policy treats 51.1% of 16-year-old males in the NLSY data set. If this magnitude of intervention was desired, universal intervention may make more practical sense than screening over half of 16-year-old males, particularly given the questionable compliance of this age group. Taken together, although our model has many shortcomings, in the very least our results raise the possibility that universal intervention is preferable to intervention that is targeted via screening. See

Cecchini et al. (2010) for a study that compares the cost-effectiveness of various obesity interventions.

6.1. Limitations of the Study

Childhood obesity screening and treatment is a prime candidate for mathematical modeling and analysis because of the impracticality of running 50-year clinical trials that link childhood BMI to adult disease. The USPSTF policy has not been previously analyzed with respect to suboptimality or supply-side feasibility (over five million children per year would be referred for treatment under the USPSTF policy), and our study is intended to inform the debate on the appropriate screening policy for childhood obesity. However, our results do not imply that the USPSTF policy should be overturned for two reasons: our model captures only some of the aspects of this very complex issue, and the lack of available data leads to some modeling assumptions that are not based on strong evidence.

Our model does not include a number of issues that—with a stronger evidence base—could conceivably provide arguments to justify the policy recommendations of the USPSTF, even if our numerical results are accurate: e.g., (i) the social and psychological effects of childhood obesity; (ii) treatment also instills other good lifestyle choices (e.g., no smoking) that provide health benefits later in life (Magnussen et al. 2011); (iii) screening earlier would prevent type 2 diabetes (the prevalence of type 2 diabetes in childhood is on the order of 0.1% and is concentrated in Native American populations and in minority children with a family history of diabetes; Fagot-Campagna et al. 2000) or the possibly irreversible effects of nonalcoholic fatty liver disease (Alisi et al. 2009) in childhood; (iv) the effect of prepregnancy weight on birth outcome and obesity risk in the next generation (this factor may lead to screening girls a few years earlier than in Figure 2, but not at age 6); or (v) the tangential health benefits accrued by other family members as a result of family involvement in intervention might be larger when the treated children are younger. Although some of these effects (e.g., issue (i)) are difficult to quantify, better data on health outcomes related to these issues would help to inform the debate and could possibly be incorporated into future models.

Aside from the omission of the issues listed previously, most of the limitations of our study are due to the paucity of data. The data include BMI evolution, the effect of treatment on BMI, and the effect of BMI on disease, which we consider in turn. The stochastic model for the evolution of the population-wide BMI pdf captures the salient features of the NLSY data and appears to be at the appropriate level of complexity for the strategic issues addressed here. The BMI

evolution data up to age 18 are plentiful but somewhat dated (and also partially self-reported, which may understate BMI levels): the NLSY data cover children born between 1970 and 1988, and the PSID data used for this analysis cover adults born in the late 1960s and early 1970s who were not children during the obesity epidemic. Childhood obesity has tripled since 1980 but has leveled off in the last decade with the exception of the far right tail of the BMI distribution for boys (Ogden et al. 2010). Hence, our results underestimate the fraction of children who would screen positive, and our analysis should be repeated when more recent data become available; nonetheless, the nature of the optimal policy depends on the temporal manner in which obese children's weight fluctuates over time in the absence of treatment, not on how many children are obese, and so our qualitative conclusions are unlikely to change if we use more recent data. In addition, although our model looks at each gender separately, the NLSY data set is not sufficiently large to look at other subgroups (e.g., racial or socioeconomic). Finally, better longitudinal data beyond age 18—both more frequent measurements and data beyond age 40—would strengthen the analysis (the trade-off curve would likely be steeper if we minimized disease prevalence at age 50 or age 60).

Turning to the effect of treatment on BMI, although the NLSY BMI data by and large follow the Markov property in the absence of treatment, our model also assumes that the Markov property holds after treatment. That is, if a child receives treatment at age t and consequently has BMI b_{t+2} (a realization of the random variable B_{t+2}) two years later, then his future BMI trajectory is statistically identical to a child who has never received treatment but who also has BMI b_{t+2} at age $t+2$. This Markov assumption can cut both ways. On one hand, data suggest that some treatment effects are transient (Oude Luttikhuis et al. 2009), and others are not (Epstein et al. 2007). On the other hand, treatment could have a long-term impact on healthy behaviors (e.g., food choices and exercise patterns). This Markov assumption deserves more attention in the future. In addition, we may be overestimating the impact of multiage policies if an individual's efficacies across different treatments are positively correlated: we treat them as independent and identically distributed, and people with lower efficacy are more apt to exceed the threshold in subsequent years. In this regard, it is worth noting that the optimal single-age policy outperforms the biennial USPSTF policy (except for diabetes in women) and is nearly optimal if obesity is measured by z-BMI.

Perhaps the most unrealistic aspect of the model formulation is that it implicitly assumes 100% treatment compliance for all children who screen positive. In practice, many families will refuse treatment for their children, even if the treatment is free. Moreover, compliance among those who initiate treatment may be lower than the compliance achieved in clinical trials. However, if compliance is independent of pretreatment BMI level and age, then the model (in particular, Equation (1)) is easily modified to incorporate imperfect compliance. Extensive numerical results in Wein et al. (2012) suggest that the relative reduction in disease prevalence (at the same treatment prevalence) under imperfect compliance is slightly less than the compliance rate times the relative reduction in the 100% compliance case, and the relative reduction in treatment prevalence (at the same disease prevalence) is 15%–30% less than it is under 100% compliance. Hence, the overall qualitative conclusion about the optimal policy versus the USPSTF policy should be unaffected by the compliance rate as long as compliance is not age dependent. As noted in the next paragraph, although the compliance rate may vary by age, there are not sufficient data to quantify this effect.

Given that it is only very recently that treatment has been shown to be moderately effective at reducing BMI (Whitlock et al. 2010), it is not surprising that the data on the effect of treatment on BMI are sparse. Indeed, although our model can accommodate efficacy that varies by gender, age, and pretreatment BMI, the existing data do not allow us to take full advantage of this. In the near term, this—along with compliance data (which may be indirectly quantified by the effect of treatment on BMI)—probably represents the most pressing data requirement. Despite this lack of data, we assume that the effect of treatment on the transformed BMI, X_t, is independent of age and pretreatment BMI level. This assumption forces the reduction in untransformed BMI due to treatment to increase with pretreatment BMI, which fits the data from the high-intensity programs in Table 1 of Whitlock et al. (2010) (see Figure 4 of the electronic companion). This assumption also leads to a larger z-BMI reduction for ages 6, 8, and 10 than for ages 12, 14, and 16, which is consistent with the studies summarized in Epstein et al. (2007). Our sensitivity analysis in §5.3 provides a crude assessment of this issue: the ratio of the absolute value of BMI reduction due to treatment for ages 6, 8, and 10 divided by the absolute value of BMI reduction due to treatment for ages 12, 14, and 16 needs to increase by 60% relative to our base-case assumptions for the optimal biennial policy to achieve the same performance as the optimal $(14, 16)$ policy. Given that adolescents are typically under less parental control than younger children and that a 60% increase does not appear to be beyond the realm of feasibility, the effect of age on treatment compliance and efficacy deserves further investigation. In contrast, our sensitivity analysis in §5.3 shows that our qualitative conclusions are

insensitive to the possibility that children above the 97th percentile of the CDC BMI distributions achieve significantly larger BMI reductions than those below the 97th percentile, or if, in our key assumption, obesity is measured by z-BMI rather than BMI.

Finally, the data on the lifetime effects of BMI on adult disease are by far the most difficult to obtain, perhaps requiring a longitudinal study of at least five decades. In the absence of additional studies on this topic, we believe that it is prudent to employ our key assumption that adult disease depends on the average lifetime obesity. Our sensitivity analysis in §5.3 on this issue shows that our key assumption is conservative with respect to the argument that the optimal policy outperforms the biennial USPSTF policy. That said, one effect that our model ignores (because we use separate data sets for ages $[2, 18]$ and $[18, 40]$) is that infants who are very small until age 2 and then experience a catch-up increase in BMI soon thereafter have an increased risk of type 2 diabetes as an adult (Eriksson et al. 2003, Bhargava et al. 2004), perhaps because of changes in metabolic factors. Although this phenomenon may be more prevalent in the developing world (indeed, our Markov test results suggest that this effect is not prevalent in the NLSY data), it does raise the possibility that more complicated policies that screen for this phenomenon (screening, although not treatment, would need to start by age 2) could enhance screening performance. We leave this as a topic for future research.

6.2. Conclusion

Although additional treatment efficacy and compliance data and more recent longitudinal BMI data would allow us to refine our analysis, and the lifetime effects of obesity on adult disease are not well understood, our results—which consequently are quite fragile—suggest that the most effective way to reduce adult disease via childhood obesity screening and treatment may be to focus on older adolescents. The fragility of our results, coupled with the fact that our model captures only some of the aspects of this complex issue, preclude us from making policy recommendations regarding childhood obesity and screening policies. Hence, although our numerical results are at odds with the USPSTF recommendations, we feel that our main contribution is not in the policy realm but in the problem framing and modeling: viewing the issue as a constrained optimization problem of minimizing adult disease prevalence subject to a treatment budget, modeling the evolution of the population-wide distribution of BMI, investigating the Markov behavior of BMI and the normality of BMI and its increments, the modeling of treatment with a discussion of the three modeling options, and the key assumption underlying the calculation of the adult disease prevalence. More specifically, we find

that—for our set of parameter values—the solution to the dynamic program, which assumes a Markov BMI process, Box–Cox normality of the BMI increments, and a separable objective function, appears to be very similar to the optimal solution for the more complex simulation model. Although our modeling choices may not all stand the test of time, they do provide a basis for discussion and a framework on which to build. Our analysis also highlights the need for treatment efficacy data that are broken down by age, gender, pretreatment BMI level and treatment history, and—more ambitiously—longitudinal data on the effect of BMI levels throughout an individual's lifetime on morbidity and mortality. Until these data become available, policy makers will need to assess various policies in the face of incomplete information.

References

Abraham S, Collins G, Nordsieck M (1971) Relationship of childhood weight status to morbidity in adults. *Health Services and Mental Health Administration Health Reports* 86:273–284.

Alisi A, Manco M, Vania A, Nobili V (2009) Pediatric nonalcoholic fatty liver disease in 2009. *J. Pediatrics* 155:469–474.

Barrett-Connor E (1989) Epidemiology, obesity, and non-insulin-dependent diabetes mellitus. *Epidemiologic Reviews* 11:172–181.

Benes VE, Shepp LA, Witsenhausen HS (1980) Some solvable stochastic control problems. *Stochastics* 4:39–83.

Berenson GE, Srinivasan SR, Bao W, Newman WP III, Tracy RE, Wattigney WA (1998) Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. *New England J. Medicine* 338:1650–1656.

Bertsekas DP (2005) Rollout algorithms for constrained dynamic programming. Technical Report LIDS-2646, MIT, Cambridge, MA. Accessed February 15, 2010, http://web.mit.edu/dimitrib/www/Rollout_Constrained.pdf.

Bhargava SK, Sachdev HS, Fall CHD, Osmond C, Lakshmy R, Barker DJP, Biswas SKD, Ramji S, Prabhakaran D, Reddy KS (2004) Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *New England J. Medicine* 350:865–875.

Box GEP, Cox DR (1964) An analysis of transformations. *J. Royal Statist. Soc. B* 26:211–252.

Bureau of Labor Statistics (2010) National longitudinal surveys: NLSY79 children and young adults. U.S. Department of Labor, Washington, DC. Accessed August 12, 2010, www.bls.gov/nls/nlsy79ch.htm.

Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D (2010) Tackling of unhealthy diets, physical inactivity, and obesity: Health effects and cost-effectiveness. *Lancet* 376:1775–1784.

Colditz GA, Willett WC, Stampfer MJ, Manson JE, Hennekens CH, Arky RA, Speizer FE (1990) Weight as a risk factor for clinical diabetes in women. *Amer. J. Epidemiology* 132:501–513.

Coleman KJ, Tiller CL, Sanchez J, Heath EM, Sy O, Milliken G, Dzewaltowski DA (2005) Prevention of the epidemic increase in child risk of overweight in low-income schools. *Arch. Pediatric Adolescent Medicine* 159:217–224.

D'Amato R, D'Aquila R, Wein LM (2000) Management of antiretroviral therapy for HIV infection: Analyzing when to change therapy. *Management Sci.* 50:1200–1213.

Epstein LH, Paluch RA, Roemmich JN, Beecher MD (2007) Family-based obesity treatment, then and now: Twenty-five years of pediatric obesity treatment. *Health Psych.* 26:381–391.

Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ (2003) Early adiposity rebound in children and risk of type 2 diabetes in adult life. *Diabetologia* 46:190–194.

Fagot-Campagna A, Pettitt DJ, Engelgau MM, Burrows NR, Geiss LS, Valdez R, Beckles GL, et al. (2000) Type 2 diabetes among North American children and adolescents: An epidemiologic review and a public health perspective. *J. Pediatrics* 136:664–672.

Finkelstein E, Trogdon J, Cohen J, Dietz W (2009) Annual medical spending attributable to obesity: Payer- and server-specific estimates. *Health Affairs* 28:w822–w831.

Franks PW, Hanson RL, Knowler WC, Moffett C, Enos G, Infante AM, Krakoff J, Looker HC (2007) Childhood predictors of young-onset type 2 diabetes. *Diabetes* 56:2964–2972.

Freedman DS, Khan LK, Dietz WH, Srinivasan SR, Berenson GS (2001) Relationship of childhood obesity to coronary heart disease risk factors in adulthood: The Bogalusa heart study. *Pediatrics* 108:712–718.

Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH (2007) Cardiovascular risk factors and excess adiposity among overweight children and adolescents: The Bogalusa heart study. *J. Pediatrics* 150:12–17.

Garber AM, Olshen RA, Zhang H, Venkatraman ES (1994) Predicting high-risk cholesterol levels. *Internat. Statist. Rev.* 62:203–228.

Goldhaber-Fiebert JD, Rubinfield RE, Bhattacharya J, Robinson TN, Wise PH (2012) The utility of childhood and adolescent obesity assessment in relation to adult health. *Medical Decision Making*, ePub ahead of print May 29, doi: 10.1177/0272989X12447240.

Gortmaker SL, Peterson K, Wiecha J, Sobol AM, Dixit S, Fox MK, Laird N (1999) Reducing obesity via a school-based interdisciplinary intervention among youth. *Arch. Pediatric Adolescent Medicine* 153:409–418.

Gunnell DJ, Frankel SJ, Nanchahal K, Peters TJ, Smith GD (1998) Childhood obesity and adult cardiovascular mortality: A 57-y follow-up study based on the Boyd Orr cohort. *Amer. J. Clinical Nutrition* 67:1111–1118.

HEALTHY Study Group (2010) A school-based intervention for diabetes risk reduction. *New England J. Medicine* 363:443–453.

Hubert HB, Feinleib M, McNamara PM, Castelli WP (1983) Obesity is an independent risk factor for cardiovascular disease: A 26-year follow-up of participants in the Framingham heart study. *Circulation* 67:968–977.

Institute for Social Research (2010) Panel study of income dynamics. Accessed August 12, 2010, http://psidonline.irs.umich.edu/.

Institute of Medicine (2005) *Preventing Childhood Obesity: Health in the Balance* (National Academies Press, Washington, DC).

Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL (2000) CDC growth charts: United States. *Advance Data* 314:1–27.

Lavieri M, Puterman M, Tyldesley S, Morris W (2012) When to treat prostate cancer patients based on their PSA dynamics. *IIE Trans. Health Care Systems* 2:62–77.

Lee CP, Chertow GM, Zenios SA (2008) Optimal initiation and management of dialysis therapy. *Oper. Res.* 56:1428–1449.

Magnussen CG, Thomson R, Cleland VJ, Ukoumunne OC, Dwyer T, Venn A (2011) Factors affecting the stability of blood lipid and lipoprotein levels from youth to adulthood. *Arch. Pediatr. Adosesc. Medicine* 165:68–76.

Mahoney LT, Burns TL, Stanford W, Thompson BH, Witt JD, Rost CA, Lauer RM (1996) Coronary risk factors measured in childhood and young adult life are associated with coronary artery calcification in young adults: The Muscatine Study. *J. Amer. College Cardiology* 27:277–284.

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed. *Frontiers in Econometrics* (Academic Press, New York), 105–142.

Mission: Readiness—Military Leaders for Kids (2010) Too fat to fight: Retired military leaders want junk food out of America's schools. Report, Mission: Readiness, Washington, DC.

Must A, Jacques PF, Dallal GE, Bajema CJ, Dietz WH (1992) Long-term morbidity and mortality of overweight adolescents: A follow-up of the Harvard Growth Study of 1922 to 1935. *New England J. Medicine* 327:1350–1355.

National Center for Health Statistics (2006) Analytic and reporting guidelines: The National Health and Nutrition Examination Survey (NHANES). Centers for Disease Control and Prevention, Hyattsville, MD.

Nemet D, Barkan S, Epstein Y, Friedland O, Kowen G, Eliakim A (2005) Short- and long-term beneficial effects of a combined dietary-behavior-physical activity intervention for the treatment of childhood obesity. *Pediatrics* 115:e443–e449.

Nihiser AJ, Lee SM, Wechsler H, McKenna M, Odom E, Reinold C, Thompson D, Grummer-Strawn L (2007) Body mass index measurement in schools. *J. School Health* 77:651–671.

Ogden CL, Carroll MD, Curtin LR, Lamb MM, Flegal KM (2010) Prevalence of high body mass index in U.S. children and adolescents, 2007–2008. *J. Amer. Medical Assoc.* 303:242–249.

O'Hagan A, Leonard T (1976) Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* 63:201–202.

Oude Luttikhuis H, Baur L, Jansen H, Shrewsbury VA, O'Malley C, Stolk RP, Summerbell CD (2009) Interventions for treating obesity in children (Review). *The Cochrane Library*, Issue 1 (John Wiley & Sons, New York), Article CD001872.

Raitakari OT, Juonala M, Kahonen M, Taittonen L, Laitinen T, Maki-Torkko N, Jarvisalo MJ, et al. (2003) Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: The cardiovascular risk in young finns study. *J. Amer. Medical Assoc.* 290:2277–2283.

Reilly JJ, Dorosty AR, Emmett PM (2000) The ALSPAC study team. Identification of the obese child: Adequacy of the body mass index for clinical practice and epidemiology. *Internat. J. Obesity* 24:1623–1627.

Reinehr T, de Sousa G, Toschke AM, Andler W (2006) Long-term follow-up of cardiovascular disease risk factors in children after an obesity intervention. *Amer. J. Clinical Nutrition* 84:490–496.

Savoye M, Shaw M, Dziura J, Tamborlane WV, Rose P, Guandalini C, Goldberg-Gell R, et al. (2007) Effects of a weight management program on body composition and metabolic parameters in overweight children. *J. Amer. Medical Assoc.* 297:2697–2704.

Shechter SM, Bailey MD, Schaefer AJ, Roberts MS (2008) The optimal time to initiate HIV therapy under ordered health states. *Oper. Res.* 56:20–33.

Singh AS, Mulder C, Twisk JW, van Mechelen W, Cinapaw MJ (2008) Tracking of childhood overweight into adulthood: A systematic review of the literature. *Obesity Rev.* 9:474–488.

U.S. Preventive Services Task Force (USPSTF) (2005) Screening and interventions for overweight in children and adolescents: Recommendation statement. *Pediatrics* 116:205–209.

U.S. Preventive Services Task Force (USPSTF) (2010) Screening for obesity in children and adolescents: U.S. Preventive Services Task Force recommendation statement. *Pediatrics* 125:361–367.

Waters E, de Silva-Sanigorski A, Hall BJ, Brown T, Campbell KJ, Gao Y, Armstrong R, Prosser L, Summerbell CD (2011) Interventions for preventing obesity in children (Review). *The Cochrane Library*, Issue 12 (John Wiley & Sons, New York), Article CD001871.

Wein LM, Yang Y, Goldhaber-Fieber JD (2012) Assessing screening policies for childhood obesity. *Obesity* 20:1437–1443.

White House Task Force on Childhood Obesity (2010) *Solving the Problem of Childhood Obesity Within a Generation* (Domestic Policy Council, Washington, DC).

Whitlock EP, O'Connor EA, Williams SB, Beil TL, Lutz KW (2010) Effectiveness of weight management interventions in children: A targeted systematic review for the USPSTF. *Pediatrics* 125:e396–e418.

Willett WC, Manson JE, Stampfer MJ, Colditz GA, Rosner B, Speizer FE, Hennekens CH (1995) Weight, weight change, and coronary heart disease in women: Risk within the "normal" weight range. *J. Amer. Medical Assoc.* 273:461–465.

Wilsgaard T, Arnesen E (2007) Body mass index and coronary heart disease risk score: The Tromso Study, 1979 to 2001. *Ann. Epidemiology* 17:100–105.