Januray 2013

Carrots, Sticks and False Carrots: How high should weight control wellness incentives be? Findings from a population-level experiment

Harald Schmidt
Lecturer, Department of Medical Ethics and Health Policy, Research Associate, Center for Health Incentives and Behavioral Economics, School of Medicine, University of Pennsylvania, schmidth@exchange.upenn.edu

Follow this and additional works at: https://uknowledge.uky.edu/frontiersinphssr

Part of the Bioethics and Medical Ethics Commons, Health Law and Policy Commons, Health Services Research Commons, and the Preventive Medicine Commons

Recommended Citation
DOI: 10.13023/FPHSSR.0201.02

This Article is brought to you for free and open access by the Center for Public Health Systems and Services Research at UKnowledge. It has been accepted for inclusion in Frontiers in Public Health Services and Systems Research by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Carrots, Sticks and False Carrots: How high should weight control wellness incentives be? Findings from a population-level experiment

ABSTRACT

Employers are increasingly using wellness incentives, including penalties for unhealthy behavior. Survey data suggests that people are willing to accept the principle of penalizing those perceived to take health risks, but the equally relevant question of the magnitude of acceptable penalties is unclear.

While the principle of penalizing overweight and obese people has some support, findings from a population-level experiment (n=1,000) suggest that the acceptable size of penalties is comparatively small, around $50: more than 10-fold below levels favored by advocates. Reward-based incentives are favored over penalty-based ones by a factor of 4. Of two different ways of framing penalty programs, poorer and higher weight groups appear to find the one that is more overtly penalizing less acceptable.

Levels of incentives matter on effectiveness as well as on ethical grounds, as it cannot be assumed that it is equally easy for all to meet health targets to secure a benefit or avoid a penalty. Programs should be designed to engage, not to frustrate those most in need of health improvement. Employee involvement in determining incentive types and levels, and explicit justification for program design can help both employees and employers to reap benefits.

Keywords
wellness incentives, fairness, equity, obesity, health promotion

Cover Page Footnote
I am grateful for discussion with David Asch, Scott Halpern, Julian Le Grand, Andrea Troxel, Kevin Volpp, and Jingsan Zhu on various aspects of designing and interpreting the experiment. I am also grateful for the insightful comments from the three anonymous reviewers, which helped to clarify important aspects that were insufficiently clear. The usual caveats apply, and in particular, all errors are mine alone. This research was supported by a grant from the National Institute on Aging, P30-AG034546. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

This Article is available in Frontiers in Public Health Services and Systems Research: https://uknowledge.uky.edu/frontiersinphssr/vol2/iss1/2
Introduction

Financial incentives may be useful tools for health promotion and cost control. Over recent years, employers’ use of wellness incentives has increased substantially, and 2010 legal reforms introduced considerably higher levels for both reward and penalty-based incentives. Survey research suggests some support for the principle of penalizing people who are perceived to be taking health risks. However, little is known about the magnitude of acceptable penalties and rewards among the public. The levels of incentives matter on effectiveness as well as fairness grounds, as it cannot be assumed that it is equally easy for all to meet targets to secure rewards or avoid penalties. A population-level online survey experiment was fielded to 1,000 US residents to explore acceptable levels of incentives and framing effects. Drawing on a typical employer-offered weight control program, reward-based incentives were favored over penalty-based ones by a factor of 4. The magnitude of acceptable penalty was around $50, 14 times lower than advocated by leading proponents. Penalty-levels in the target population were lower still. Of two different ways of framing penalty programs, poorer and higher weight groups appear to find the one that is more overtly penalizing less acceptable.

Section 2705 of the Affordable Care Act has increased the permissible levels for wellness incentives tied to health outcomes. While previously an amount of 20% of the cost of coverage was permissible to be used as incentives, from 2014, 30-50% may be put at stake. According to a recent survey, employer use has more than doubled since 2009: then, 36% of large employers used wellness incentives, and around 80% are expected to so in 2012. The use of penalties saw a more than four-fold increase over this period: 8% in 2009, and 38% in 2012. These developments make an exploration of consumers’ perspectives urgent.

Methods

In this pilot study, an experiment as part of an attitudinal 14-item cross-sectional online-survey was fielded to working-age US residents (20-65 years) via the platform SurveyMonkey for 10 days in September/October 2011. Respondents were members of a large opt-in ‘river-sampled’ market research panel managed by USamp. Census-adjusted quota sampling by income was performed (12 income levels in roughly $10,000 intervals), and the survey was stopped when n=1,000 responses were received.

USamp panelists received a generic email with an invitation to a health-related survey and a link to the study-website, which provided more detailed information about aims, content and duration. Nine percent of those who received the generic Usamp invitation completed the survey, a typical response rate for this kind of panel. Of those who started the survey, 97% completed it. The sample’s mean age was 40 years (SD=13), and gender proportions were 50% male/female. The mean difference regarding income, insurance status, health and body mass index (BMI) levels in comparison with census data was 5.2% (see also Fig. 1). Data were examined using one-way non-parametric analysis (Wilcoxon/Kruskal-Wallis test). Co-variates included age, gender, income, health, BMI.

Results

The experiment entailed 3 scenarios with different incentive structures that were presented in randomized order. The scenario stem was that an employer offers employees the opportunity to be weighed at 6-month intervals:

- In the “carrot” (i.e.: net-reduction) scenario, employees with normal BMI on 2 consecutive weigh-ins receive an insurance premium rebate.
- In the “stick” (i.e: net-increase) scenario, employees who fail to meet the normal BMI target face a surcharge.
- To test framing effects, the third scenario was a “false-carrot” format: insurance contributions are increased for all employees at the beginning of the plan year by a certain amount, and the increase is reimbursed if the target is met. This framing might be perceived as a reduction, but - just as in the “stick” scenario - results in a surcharge for not meeting standards.

In all 3 scenarios respondents were asked to determine appropriate levels by entering amounts between $0 (to indicate disagreement) and $2500 (the cap representing the approximate maximum amount in 2014, based on the average cost of coverage).

Fig 1 shows means and medians for all amounts including $0, and for amounts larger than $1. Due to the non-normal distribution, the median of all amounts is the most appropriate as a measure of general approval, i.e. $200 for
“carrot”, and $50 for both “stick” and “false carrot” (interquartile range: “carrot”: $51-458; both remaining: $0-$200). While none of the scenarios had more than n=16 missing responses, around 40% of respondents entered $0 in the penalty-based scenarios, indicating twice higher levels of disapproval (18% entered $0 for “carrot”, 37% for “stick” and 42% for “false carrot”). Conversely, around 60% agreed with some kind of penalty for those not meeting BMI targets.

Subgroup analysis revealed significant differences across age, income, health and BMI groups, although not for gender. Fig. 2 shows that higher weight respondents generally set lower levels in all scenarios, except for “carrot”, where both the overweight and obese opted for $150 and those with normal weight for $200. While the median of the normal weight group for “stick” and “false carrot” was

### Key characteristics of the sample in comparison with the US population

<table>
<thead>
<tr>
<th></th>
<th>US (%)</th>
<th>Sample (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (&gt;200% of median gross household income [HHI])</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Middle (60–200% HHI)</td>
<td>49</td>
<td>56</td>
</tr>
<tr>
<td>Low (&lt;60% HHI)</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td><strong>Weight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal (BMI 18.5-25)</td>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>Overweight (BMI 25-30)</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Obese (BMI &gt;30)</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td><strong>Health</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent/very good</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>Good</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>Fair/poor</td>
<td>14</td>
<td>16</td>
</tr>
</tbody>
</table>

identical at $100, the obese and overweight set lower amounts for the former 
(overweight: $28 vs. $50; obese $0 vs. $10).

A very similar pattern was observed with regard to the three income 
groups. The best-off and the middle group entered $200 for “carrot”, the low-
income group $150. For “stick”, the high-income group converged around $100. 
The middle income group set $50, and the low-income group $20. For “false 
carrot”, low and middle income groups both opted for $50. The high-income 
group doubled this again, at $100.

![Fig. 2: Levels (in $) for 3 types of incentives by BMI and income (Mdn)](image)

<table>
<thead>
<tr>
<th></th>
<th>Carrot</th>
<th>False carrot</th>
<th>Stick</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>200</td>
<td>100–500</td>
<td>100</td>
</tr>
<tr>
<td>Overweight</td>
<td>150</td>
<td>50–300</td>
<td>50</td>
</tr>
<tr>
<td>Obese</td>
<td>150</td>
<td>8–272</td>
<td>10</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>H(2)=19.81, p&lt;.0001</td>
<td>H(2)=20.28, p&lt;.0001</td>
<td>H(2)=34.27, p&lt;.0001</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>200</td>
<td>100–500</td>
<td>100</td>
</tr>
<tr>
<td>Middle</td>
<td>200</td>
<td>100–500</td>
<td>50</td>
</tr>
<tr>
<td>Low</td>
<td>150</td>
<td>15–295</td>
<td>50</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>H(2)=20.66, p&lt;.0001</td>
<td>H(2)=10.80, p=0.045</td>
<td>H(2)=12.72, p&lt;.0022</td>
</tr>
</tbody>
</table>

Regarding the framing of penalties as “sticks” or “false carrots”, the best-off 
weight and income groups set identical amounts. This is plausible as the financial 
burden for not meeting the target is exactly the same. However, it is noteworthy 
that the lowest income group, the overweight, and the obese set lower levels for 
“stick” than for “carrot”. This may indicate that more overt penalising by singling 
out is less welcome among poorer and higher weight groups.
Limitations

Although quota sampling enabled close approximation of census data in all relevant demographics, no other standardizations were performed. External validity is therefore limited, and the study does not permit inferences to the US population as a whole. Respondents were asked to set acceptable levels in general terms: the set amounts are therefore likely to mix assumptions about fairness and effectiveness. For technical reasons, all levels entered were analyzed, and not only those for the first of the randomized scenarios (which would have reduced the impact of possible ordering effects). In a preceding vignette, respondents rated the acceptability of a similar weight control scenario in which $150 was offered, with likely anchoring effects. Further research, extending also to other relevant health domains, such as smoking, should therefore respond to these constraints to generate more robust data (in progress).

Implications

To my knowledge, this is the first empirical study to explore the acceptability of different ways of framing weight control incentives in a large-scale experimental setting. “Carrots” were clearly preferred over “sticks”. In contrast to the preferences of advocates of increasing the legal limits of incentives, there was little support for large penalties in any of the strata. Opposition was strongest among low-income groups, the overweight, and the obese. The findings can suggest that where larger penalties are used, frustration and perhaps even pushback is possible. Care is required to ensure that employees do not perceive any form of incentive program merely as unfair cost-shifting, and reject the approach as a whole. Genuine employee engagement can help ascertain employees’ attitudes on levels and types of incentives early on in the planning process. Explicit justification should be offered regarding program design and rationales, and the basis on which targets are deemed achievable for employees.2 Wellness incentives are likely to be most effective when the interests of employees, employers and policy makers are aligned.
Summary Box:
What is already known on this topic?
Employers are increasingly using wellness incentives, including penalties for unhealthy behavior. Survey data suggests that people are willing to accept the principle of penalizing those perceived to take health risks, but the equally relevant question of the magnitude of acceptable penalties is unclear.

What is added by this report?
While the principle of penalizing overweight and obese people has some support, findings from a population-level experiment (n=1,000) suggest that the acceptable size of penalties is comparatively small, around $50: more than 10-fold below levels favored by advocates. Reward-based incentives are favored over penalty-based ones by a factor of 4. Of two different ways of framing penalty programs, poorer and higher weight groups appear to find the one that is more overtly penalizing less acceptable.

What are the implications for public health practice/policy/research?
Levels of incentives matter on effectiveness as well as on ethical grounds, as it cannot be assumed that it is equally easy for all to meet health targets to secure a benefit or avoid a penalty. Programs should be designed to engage, not to frustrate those most in need of health improvement. Employee involvement in determining incentive types and levels, and explicit justification for program design can help both employees and employers to reap benefits.