Medical Decision Making

# Levelling up: treating uptake as endogenous may increase the value of screening programmes

| | |
|---|---|
| Journal: | *Medical Decision Making* |
| Manuscript ID | MDM-24-043.R3 |
| Manuscript Type: | Original Research Article |
| APPLICATION AREAS: | ONCOLOGY, Cancer Prevention < ONCOLOGY, Preventive Medicine--Screening < PUBLIC HEALTH |
| DETAILED METHODOLOGY: | Diagnostic Test Evaluation < CLINICAL RESEARCH METHODOLOGY, Provider Decision Making < DECISION AIDS--TOOLS, ECONOMICS (HEALTH) |
| | |

# Levelling up: treating uptake as endogenous may increase the value of screening programmes

**Jose A. Robles-Zurita, PhD**
*(Corresponding email address:* jarobzur@uma.es)
*Department of Applied Economics (Statistics and Econometrics), University of Málaga, Spain*
*Health Economics and Health Technology Assessment, University of Glasgow, UK*

**Neil Hawkins, PhD**
*Health Economics and Health Technology Assessment, University of Glasgow, UK*

**Janet Bouttell, PhD**
*Health Economics and Health Technology Assessment, University of Glasgow, UK*

***Background***. We aimed at illustrating that health economists should consider individual heterogeneity when solving the problem of finding the optimal combination of sensitivity and specificity that maximises average health utility of a target population in a screening programme. ***Methods***. A theoretical framework compares the solution under standard economics of diagnoses to the optimal combination under an endogenous uptake analysis, where screening participation is given by heterogenous health preferences. An applied example used calibrated parameters with real data from the bowel cancer screening programme in the UK. Scenario analyses show how the results would change with parameters values, if disease risk and health utilities are not independent, and if screening uptake was not completely determined by health preferences. ***Results***. A general theoretical result states that the endogenous uptake analysis leads to a weakly higher true and false positive rate than would be optimal under the standard approach. In the same way, the endogenous solution would lead to a lower uptake rate. The base case scenario of the applied example illustrates that a screening programme using the endogenous solution would generate 21.1% more quality adjusted life years (QALYs) than when using the standard solution. The scenario analyses show when the endogenous analysis is most valued; and that the general result applies for a wide range of situations when theoretical assumptions are relaxed. ***Limitations***. The results obtained are valid under the assumptions made. Analysts should evaluate if those could hold in the applied screening context. ***Conclusions***. Individual heterogeneity and uptake decisions are relevant factors to consider in the problem of finding an optimal combination of sensitivity and specificity for a screening test.



**Highlights**

- The value of screening programmes can be higher if heterogeneity of preferences in the target population is considered.
- The optimal operation of a screening test depends on health utilities of the target population and on the heterogeneity of these health utilities.
- Under heterogeneity of health utilities, the optimal operation of a screening test does not maximise screening uptake.
- A general theoretical result states that the endogenous uptake analysis leads to a weakly higher true and false positive rate than would be optimal under a standard approach; this is true for a wide range of situations.

**Keywords**

Screening test; sensitivity; specificity; screening; optimal test; endogenous uptake

## Introduction

The health economics literature has pointed to the importance of accounting for individual preferences and uptake when evaluating health interventions; specifically, the importance of modelling uptake as endogenous, i.e. being a function of the characteristics of the interventions.[1] Although this is an economic problem affecting many types of health policy, we are interested in studying the case of screening tests. The screening test accuracy affects the benefit that it provides. For example, a test can improve the health of the population by identifying individuals ('true positives') who may benefit from medical intervention. However, this benefit needs to be balanced against the disbenefit (in terms of cost and potential negative health effects of unnecessary interventions) in patients who do not ultimately benefit from the intervention ('false positives'). The balance of these outcomes will depend on the performance of the test in terms of sensitivity and specificity; i.e., the percentage of sick and healthy people, respectively, that are correctly diagnosed. For individuals, this balance will also depend on their own preferences with respect to the benefits of treatment if they happen to be a true positive and disbenefits if they happen to be a false positive, and their prior belief that their participation in screening will lead to true or false positive results. Individuals' decision to participate in a screening programme will depend on their personal judgements regarding the net benefit of participation. In turn this will determine both the uptake of the screening and the valuation of a screening programme amongst those who participate. Often, there is a trade-off between sensitivity and specificity, in such a way that we can only increase sensitivity at the expense of lower levels of specificity. The choice of where to operate a screening test will impact an individual's valuation of a screening programme and their participation decisions (i.e. the uptake rate).

To our knowledge, the analysis of medical tests has never considered heterogeneity in individual preferences and its impact on uptake. For example, the "standard" economics of diagnosis shows how to choose the optimal combination of sensitivity and specificity for an individual patient rather than for a heterogeneous population.[2,3] However, a screening test performance could affect individuals' decisions about whether to participate in a screening programme. For example, there is evidence that changes in the sensitivity and specificity of a test will affect willingness to participate in cancer screening.[4] The problem of uptake endogeneity is especially relevant in the context of screening programmes where a target

population is offered a test and each individual person makes the final participation decision. In fact, uptake rates in screening programmes are far from 100% in the UK, for example about 70% in cervical and breast cancer screening, and even less than 50% for some socioeconomic subgroups in colorectal cancer screening. [5-7] This evidence suggests that a considerable proportion of the eligible population may regard the expected benefit of screening as negative and thus decline to participate.

The main purpose of this paper is to show that endogeneity is a relevant issue affecting the optimal design of screening programmes. We compare the optimal combination of sensitivity and specificity derived using standard economics of diagnostics to those derived in an endogenous uptake analysis taking account of individual heterogeneity. Section 2 presents a theoretical framework where assumptions, about the screening programme and individual decisions, are intended to serve our purpose while keeping the analysis simple. Section 3 illustrates the theoretical results with an example where parameters are calibrated using real data from the faecal immunochemical test (FIT) used in the Bowel Cancer Screening Programme (BCSP) administered by NHS-England. Section 4 is the discussion.

# Theoretical framework

## Screening test

A screening test will be defined by the true positive rate (TPR) or sensitivity, the probability of a positive result for the group of sick individuals, and the false positive rate (FPR) or 1-specificity, the probability of a positive result for the group of healthy people. We will study the case in which a test could achieve different levels of TPR and FPR given by a receiver operating characteristic (ROC) curve as in Figure 1. The points in the ROC curve represent the choice set of the economic problem analysed in this study.

## Optimal test under "standard" economics of diagnosis

Following the seminal paper by Phelps and Mushlin (1988) we can describe the different paths that can occur and the corresponding probabilities.[3] The final states after undertaking a diagnostic test could be: a) treating a sick person, if the result is a true positive; b) not treating a sick person, if the result is a false negative; c) treating a healthy person, if the result is a false positive, and; d) not treating a healthy person, if the result is a true negative. The probability

of each of those states will depend on the prevalence of the disease ($p$) in the population group and the level of TPR and FPR. The health utility (HU) of a person in the target population will depend on the final state:

a) Sick person treated: $HU_{st}$.
b) Sick person not treated: $HU_{sn}$.
c) Healthy person treated: $HU_{ht}$.
d) Healthy person not treated: $HU_{hn}$.

The expected HU of a person undergoing the test will be expressed as:

$$HU = p\{TPR * HU_{st} + (1 - TPR)HU_{sn}\} + (1 - p)\{FPR * HU_{ht} + (1 - FPR)HU_{hn}\}, \quad (1)$$

We are then interested in solving the economic problem of finding the optimal combination of TPR and FPR subject to a ROC curve. The point of maximum expected HU is the one in the ROC curve with slope equal to the HU isoquant's gradient. The slope of the isoquant is computed by applying the full differential of equation (1):

$$S = \frac{dTRP}{dFPR} = \frac{-(1-p)(\Delta \mathrm{HU}_h)}{p(\Delta \mathrm{HU}_s)}. \quad (2)$$

Where $\Delta HU_h = HU_{ht} - HU_{hn} < 0$ is the decrement in health utility for a healthy person being treated. Also, $\Delta HU_s = HU_{st} - HU_{sn} > 0$ is the incremental health utility of being treated for a sick patient

Expression (2) has been acknowledged in the literature. [2,3,8,9] That work considered costs as well as health effects in a cost-effectiveness analysis framework. In this paper, in order to focus on uptake endogeneity, we will assume that the objective of the decision maker is to maximise the health utility of a given population.

## Screening test with endogenous uptake

In this section, we will introduce the problem of deciding the optimal operation of a test used for a screening programme. Under this scenario it is usual to have a health provider, e.g., public national health system, offering a test to a target population. The screening test will operate at one certain combination of sensitivity/specificity chosen along the ROC curve for

the whole target population. Everyone in that population decides whether to participate in the screening programme. Therefore, the main modification of the health economic problem, compared to the standard analysis in section 2.2, is allowing individuals to decide whether to undergo the screening test based on its characteristics and their own preferences. Given heterogeneity in the target population (e.g. variability in health utilities or disease risk), only a percentage of people will find the test to be beneficial and, therefore, will decide to participate in the screening programme. If someone decides to undergo the test, treatment will depend on the screening result, positive or negative. On the contrary, if a person decides not to undergo the test she will not receive treatment. Therefore, a person cannot receive treatment directly; they need to participate in the screening first. This assumption is consistent with the Bowel Cancer Screening Programme (BCSP) in NHS-England, where the target population (asymptomatic population) are not offered treatment unless they test positive.

The individual utility derived from undergoing the test is affected by its accuracy (i.e., sensitivity and specificity) in such a way that uptake is endogenous to the choice of the pair (TPR, FPR) provided by the national health system. The immediate consequence for the medical decision problem is that we should consider plausible changes in uptake when searching for the optimal sensitivity and specificity of a screening test.

**Optimal combination of TPR and FPR for a screening test**

In a screening programme, where individuals choose their participation, equation (1) would not be appropriate to describe the relationship between average health utility of the target population and accuracy of the test. Instead, consider that $u^i$ is a binary indicator that takes value 1 if subject $i$ participates and 0 otherwise. Therefore, under a screening programme, the expected health utility of this individual would be:

$$HU^i_{Test} = p^i\left[TPR \times HU^i_{st} + (1 - TPR)HU^i_{sn}\right] + \left(1 - p^i\right)\left[FPR \times HU^i_{ht} + (1 - FPR)HU^i_{hn}\right], \quad (3)$$

or

$$HU^i_{No\ test} = p^i HU^i_{sn} + \left(1 - p^i\right)HU^i_{hn}, \quad (4)$$

if $u^i = 1$ or $u^i = 0$ respectively. Subjects are assumed heterogenous in the utility derived from each health state ($HU^i_{st}, HU^i_{sn}, HU^i_{ht}\ and\ HU^i_{hn}$) and in the disease risk ($p^i$). Finally, the

average health utility provided by a screening programme in an heterogenous target population of *N* subjects is given by:

$$HU_{screen} = \frac{\sum_{i=1}^{N}\left[u^i \times HU^i_{Test} + (1-u^i) \times HU^i_{No\ test}\right]}{N}, \quad (5)$$

The optimal screening test will be the combination of TPR and FPR along the ROC curve that maximises expression (5). In this problem, changes in accuracy of the screening test will affect $HU_{screen}$ in two dimensions: a) the utility derived from undergoing the test ($HU^i_{Test}$), and; b) uptake ($u^i$), since the appeal of the test will be affected. Therefore, the optimal solution will depend on the specification of uptake behaviour. In section 2.3.2, uptake behaviour is assumed to be determined by individual expected health utility; then section 2.3.3 proposes two alternatives to relax this assumption.

**Uptake based on expected health utility**

In practice, individuals' preferences for a screening test could be affected by many factors, for example: time required; own experience; beliefs about their health risks; relatives' or friends' disease history; available information; confidentiality; health and cost consequences; risks of the medical test; test setting.[10,11] To account for the endogeneity of the health economic problem here proposed, we are interested in modelling preferences regarding the accuracy of the test (sensitivity and specificity). More specifically, we will assume that individuals' preferences are given by the expected health utility. A person $i$ will undergo the screening test if and only if the test's expected health utility is higher than the expected health utility of not participating in the screening programme:

$$u^i = 1 \Leftrightarrow HU^i_{Test} > HU^i_{No\ test}. \quad (6)$$

By substituting expressions (3) and (4) in expression (6) and rearranging we obtain the next equivalent participation condition:

$$u^i = 1 \Leftrightarrow S^i = \frac{-(1-p^i)(\Delta HU^i_h)}{p^i(\Delta HU^i_s)} < R = \frac{TPR}{FPR}. \quad (7)$$

Where $\Delta HU^i_s = HU^i_{st} - HU^i_{sn} > 0$ and $\Delta HU^i_h = HU^i_{ht} - HU^i_{hn} < 0$.

Expression (7) says an individual $i$ will undergo the screening test if (and only if) the ratio of the true and false positive rate $R$ is higher than the gradient of her HU isoquant $S^i$; the existence of a participation condition has also been acknowledged previously.[12] Given

heterogeneity in health utilities and disease risk, individuals will differ in $S^i$ and in uptake behaviour. It will be assumed that the decision maker has knowledge of the aggregate distribution of the health utilities and the disease risk (and hence of $S^i$), however it does not have information of individual subjects. In this analysis, all members of the target population will have the same information, and therefore the same beliefs, about the accuracy of the test (TPR and FPR). This information is assumed to be provided by the health authorities offering the screening test.

Expression (7) implies that uptake behaviour is endogenous to the decision problem of choosing the optimal pair (FPR, TPR). Imagine that the heterogeneity in $S^i$ can be well approximated by a continuous probability distribution, then Figure 2 illustrates how changes in R will affect uptake. The shaded area in panel A represents the uptake rate given by the probability of $S^i$ being below the level $R^0$, $P(S^i < \mathrm{R}^0)$. Notice that an improvement in the test accuracy (e.g. lower FPR or higher TPR) will change the ratio to $R^1 > R^0$ in panel B and the uptake rate will increase, represented by a bigger shaded area in panel B, i.e. $P(S^i < R^1) > P(S^i < \mathrm{R}^0)$. Also notice that, participating individuals in case of $R^1$ will have a higher slope $S^i$ on average, i.e., $E[S^i | S^i < R^1] > E[S^i | S^i < R^0]$.

Further implications of this model are:

a) If treating sick patients is always positive, then a sufficiently accurate test could achieve participation of 100% of the target population. If $HU^i_{st} > HU^i_{sn}$, for all subjects, a sufficient condition for 100% uptake is TPR=1 and FPR=0, i.e. $R = \infty$. In practice it is rare for a test to achieve perfect accuracy, and the effect of treatment could be negative for some patients under some circumstances. Therefore, 100% uptake is not a realistic expectation for any test available.

b) In general, the combination of TPR/FPR which maximises uptake will not be the same combination that maximises expression (5). Imagine we are going to choose between two accuracy points (A or B) in the scenario of Figure 3 where the target population is characterised by two types of individuals (type 1 in grey and type 2 in black) with preferences given by the indifference curves, where the dashed lines represent those combinations with the same utility as no participation (TPR=0 and FPR=0). The utility of type 1 individuals is higher at point A than at point B. On the other hand, 100%

uptake will only be achieved under accuracy B, where the ratio $R = \frac{TRP}{FPR}$ is higher than the slope of the HU isoquant (indifference curve) for both types of individuals. Given that we are interested in maximising the average utility of the population as in expression (5), point B will not be optimal if the number of type 1 individuals is sufficiently large so that their utility gains outbalance the losses for type 2 individuals when choosing A rather than B. In that case, the optimal solution will be A rather than B even if this is at the expense of having type 2 individuals not willing to participate in the screening.

**Alternative specifications of uptake behaviour: participation utility and random uptake**

Assuming that participation in a screening programme is fully determined by the expected health utility could be unrealistic in some contexts. In this section, we present two alternative specifications for uptake behaviour. First, we will introduce the concept of participation utility, the fact that subjects will obtain a negative utility just for participating (e.g. time and effort invested in doing the test). Secondly, we will introduce random uptake by assuming that a proportion of the target population ignores their expected health utilities and choose to participate randomly.

*Uptake accounting for participation utility*

Uptake will be given by the next participation condition:

$$u^i = 1 \Leftrightarrow HU^i_{Test} + PU^i > HU^i_{No\ test}, \quad (8)$$

Where $PU^i$ is the participation utility for subject $i$. We will anticipate that $PU^i \leq 0, \forall i$, therefore expression (8) is a more restrictive condition than expression (7): i.e. uptake is expected to be lower. Notice that $PU^i$ could be driven by any non-health outcome that is affected by undergoing the screening test (e.g. monetary costs, time spent, physical and cognitive effort). Expression (8) assumes that subjects integrate, by addition, the expected health utility and the participation utility in a *general utility function* (of health and non-health outcomes) that explains uptake behaviour.

*Random uptake*

Under this assumption, a proportion ($m$) of subjects will participate in the screening programme following expression (7), i.e. they maximise their expected health utility.

However, the remaining subjects (representing a proportion $1-m$), will participate randomly as:

$$u^i = 1 \Leftrightarrow Random^i = 1, \quad (9)$$

Where, for each subject, $Random^i \sim Bernoulli(u^B)$ is an identically and independently distributed variable that takes value 1 with probability $u^B$ and 0 with probability $1-u^B$. Therefore, the screening participation rate for random uptake individuals equals $u^B$ and is exogenously determined. Consequently, if all subjects in the population choose randomly ( $m = 0$), the participation rate will be exogenous to the solution of maximising expression (5). On the contrary, if there were a non-zero proportion of health utility maximisers ($1 \geq m > 0$) the endogeneity problem will exist, at least to some extent.

## A general result on the endogenous vs. standard solution

Two alternative solutions to the problem of choosing TPR and FPR, along a given ROC curve, for a screening programme are compared in this analysis:

a) The "standard approach". This solution will be obtained by maximising expression (1) where the health states utilities and the disease risk are fixed parameters computed as the average for the target population: $p = \frac{\sum_{i=1}^{N} p^i}{N}$; $HU_{st} = \frac{\sum_{i=1}^{N} HU_{st}^i}{N}$; $HU_{sn} = \frac{\sum_{i=1}^{N} HU_{sn}^i}{N}$; $HU_{ht} = \frac{\sum_{i=1}^{N} HU_{ht}^i}{N}$; $HU_{hn} = \frac{\sum_{i=1}^{N} HU_{hn}^i}{N}$.

b) The "endogenous solution". This solution would entail maximising expression (5).

**Proposition.** Given a ROC curve going from point (0,0) to point (1,1), increasing and concave, uptake behaviour determined by expected health utility as in expression (7), and independence between the distribution of the disease risk and the distributions of the health state utilities, then the endogenous solution is equal or above (weakly higher TPR and FPR than) the standard solution.

Let ($TPR^S$, $FPR^S$) be the optimal combination of sensitivity and specificity under the standard approach, i.e. the maximum for the function given by expression (1). For simplicity we will use only $TPR$ to refer to one point in the ROC curve, making it implicit that each $TPR$ value combines with only one specific $FPR$ value. Now let $HU(TPR^S)$ be the value achieved

according to expression (1) at the optimal solution. Imagine in a group A we have those patients who participate under $TPR^S$, and in group B there are those who do not participate. Formally, in terms of the participation condition (7), $A = \left\{ i \,\middle|\, S^i < R^S = \frac{TPR^S}{FPR^S} \right\}$ and $\mathrm{B} = \left\{ i \,\middle|\, S^i \geq R^S = \frac{TPR^S}{FPR^S} \right\}$.

Given that the distribution of the disease risk is independent of the health states utilities, we can write $HU(TPR^S)$ as the sum of the expected health utilities of the test for the two population groups divided by *N*:

$$HU(TPR^S) = \frac{\sum_{i \in A} HU^i_{Test}(TPR^S) + \sum_{i \in B} HU^i_{Test}(TPR^S)}{N}, \quad (10)$$

Notice that the expected health utility of the test at $TPR^S$ for subjects in group B is lower or equal to the health utility of no test:

$$HU^i_{No\ test} \geq HU^i_{Test}(TPR^S), \forall i \in B, \quad (11)$$

If we assume that the ROC curve goes from the point (0,0) to the point (1,1), is increasing and concave, then those patients in group B would obtain a higher health utility from participation if and only if the TPR is reduced (along the ROC curve), so that $R = \frac{TRP}{FPR}$ is increased, and they will benefit from further reductions down to an optimal point. The level of this optimal TPR will be different for each patient due to heterogeneity. Imagine that we reduce the sensitivity of the test down to $TPR^{"} < TPR^S$ (notice that this is a move along the ROC curve so that FPR is also reduced at the same time). In the new situation the value of expression (1) will be:

$$HU(TPR^{"}) = \frac{\sum_{i \in A} HU^i_{Test}(TPR^{"}) + \sum_{i \in B} HU^i_{Test}(TPR^{"})}{N}, \quad (12)$$

and the next will hold:

$$HU^i_{Test}(TPR^{"}) > HU^i_{Test}(TPR^S)\ \forall i \in B, \quad (13)$$

Notice that because $TPR^S$ is the optimal solution under the standard approach, we have:

$$0 > HU(TPR^{"}) - HU(TPR^S). \quad (14)$$

Now we need to prove that under the endogenous approach, when we maximise expression (5), $TPR^{"}$ will always be worse than $TPR^S$. Let's first specify the value of expression (5) given the uptake decisions of individuals in group A and B under $TPR^S$ and $TPR^{"}$:

$$HU_{screen}(TPR^S) = \frac{\sum_{i\in A} HU^i_{Test}(TPR^S) + \sum_{i\in B} HU^i_{No\ test}}{N}, \quad (15)$$

$$HU_{screen}(TPR^{"}) == \frac{\sum_{i\in A} HU^i_{Test}(TPR^{"}) + \sum_{i\in B_1} HU^i_{No\ test} + \sum_{i\in B_2} HU^i_{Test}(TPR^{"})}{N}, \quad (16)$$

Notice that subjects in group A will obtain the health utility of the test ($HU^i_{Test}$) because they will participate in the screening programme both at $TPR^S$ and at $TPR^{"}$. Also, in equation (15) the health utility for individuals in group B will be $HU^i_{No\ test}$ because those subjects do not participate in the screening at $TPR^S$. However, when the accuracy of the test is changed to $TPR^{"}$ two mutually exclusive and complementary groups will be differentiated within group B: $B_1$ and $B_2$. Group $\mathrm{B}_1 = \left\{ i \,\middle|\, S^i \geq R^{"} = \frac{TPR^{"}}{FPR^{"}} \right\}$ will refer to those patients that will not undergo the screening test at $TPR^{"}$ and therefore will obtain $HU^i_{No\ test}$. Group $\mathrm{B}_2 = \left\{ i \,\middle|\, S^i < R^{"} = \frac{TPR^{"}}{FPR^{"}} \right\}$ will refer to subjects that will change their behaviour and will undergo the test at the new level $TPR^{"}$; they will obtain $HU^i_{Test}(TPR^{"})$.

For convenience let's use the next two expressions:

$$\beta = HU_{screen}(TPR^S) - HU(TPR^S) = \frac{\sum_{i\in B}\left[HU^i_{No\ test} - HU^i_{Test}(TPR^S)\right]}{N}, \quad (17)$$

$$\alpha = HU_{screen}(TPR^{"}) - HU(TPR^{"}) = \frac{\sum_{i\in B_1}\left[HU^i_{No\ test} - HU^i_{Test}(TPR^{"})\right]}{N}. \quad (18)$$

And now the subtraction:

$$\beta - \alpha = \frac{\sum_{i\in B_2}\left[HU^i_{No\ test} - HU^i_{Test}(TPR^S)\right]}{N} + \frac{\sum_{i\in B_1}\left[HU^i_{Test}(TPR^{"}) - HU^i_{Test}(TPR^S)\right]}{N} \quad (19)$$

Given expression (11) the first summand is higher or equal to 0, and given expression (13) the second summand is strictly positive. So it happens that $(\beta - \alpha) > 0$. Then, from expression (14):

$$0 > HU(TPR^{"}) - HU(TPR^{S}) > HU(TPR^{"}) - HU(TPR^{S}) - (\beta - \alpha)$$
$$== HU_{screen}(TPR^{"}) - HU_{screen}(TPR^{S}), \quad (20)$$

i.e., under the endogenous approach any sensitivity of the test $TPR^{"}$, below the solution given by the standard approach $TPR^{S}$, will be less desirable. In other words, the optimal solution under the endogenous approach is equal or above the solution given by the standard approach■

A corollary of this result is that the participation rate under the endogenous solution will be equal or lower than under the standard approach. This is given by the fact that a higher TPR in the ROC curve implies a lower ratio $R = \frac{TPR}{FPR}$; therefore, fewer individuals will meet the participation condition in expression (7).

In the next section, an applied example illustrates this theoretical result and explores the impact of the relaxation of some assumptions in the proposition. In particular: a) when the disease risk and the health states utilities are not independent; b) when uptake behaviour is affected by utility participation as in expression (8), and; c) when a percentage of subjects choose randomly as in expression (9).

## An example in bowel cancer screening

In this example, we will analyse the case of the faecal immunochemical test (FIT) used in the Bowel Cancer Screening Programme (BCSP) publicly funded by NHS-England. Although we have used reasonable care to use appropriate estimates from the literature and publicly available information, the process of parameter calibration is based on some restricted assumptions that prevent us from making any claims about the BCSP. The main purpose is to demonstrate our theoretical findings and the impact of accounting for endogenous uptake in a situation based on real-world data.

We will analyse the decision problem of choosing the optimal combination of TPR and FPR at which the screening test will be operating; where the objective will be maximising the average health utility of the target population. Model parameters will be calibrated from population data, information on screening performance and previous studies. Two solutions will be obtained numerically: a) one based on maximisation of expression (1), called the “standard

approach", and; b) the "endogenous solution" based on maximising expression (5). Optimal accuracy, uptake rates and average health utility of target population will be compared for the two approaches. A base case analysis will be performed using the calibrated parameters under the assumptions of the proposition in section 2.4. Scenario analyses will illustrate how the results will change with parameters values, and under specific alternative assumptions conveniently chosen to show discrepancies with base case analysis.

## FIT in the bowel screening programme.

In England, at the time this analysis was carried out, the whole population aged 60 to 74 was invited to participate in the screening programme by using a FIT kit sent to their homes. If a person decides to participate, they will send a stool sample following the instructions. The diagnostic variable for FIT is occult (hidden) blood in the stool which is associated with cancer and advanced adenoma (small growth that may convert into cancer). In the English bowel screening programme, if the amount of occult blood is higher than a specific diagnostic threshold, measured as micrograms of occult blood per 1 gram of faeces (μg Hb per $g^{-1}$ faeces), the person is diagnosed positive, and a colonoscopy is offered. On the other hand, if the amount of occult blood is lower than the diagnostic threshold, the individual is diagnosed negative, and no further investigation is offered.

### Calibration of parameters for the base case analysis

The parameters for the base case analysis are reported in Table 1. The supplementary material gives details of the sources of parameter estimates including detail of the calibration of the ROC curve, prevalence of bowel cancer or advanced adenoma, and health utilities.

### The ROC curve

The ROC curve for FIT will be given by the combinations of sensitivity and specificity that can be chosen by changing the diagnostic threshold. The ROC curve that we will use here will be based on a binormal model fitting the empirical curve estimated for England.[13,14] The mean occult blood for the healthy and sick population was obtained from the median reported. [13] Standard deviation of the biomarker was chosen to minimise the sum of the squared errors between the fitted and empirical curve. The distribution of the diagnostic variable was fitted

to $\theta_s \sim N(96.9, 63)$ and $\theta_h \sim N(53.3, 63)$, for the sick and healthy population, respectively (see the corresponding fitted ROC curve in the supplementary materials).

**Prevalence**

The prevalence of bowel cancer or advanced adenoma was calibrated to be consistent with the NHS-England data about percentage of people obtaining a true positive result from the FIT at the given TPR estimated for the diagnostic threshold currently set at 120 µg Hb per $g^{-1}$ faeces.[15,16] Prevalence was calibrated at 2.85% of the target population being disease positive (having either bowel cancer or advanced adenoma).

**Health utilities**

Quality adjusted life years (QALYs) were used for the health utilities outcomes of the analysis, considering quantity and quality of expected remaining life years. The first utility parameter calibrated was $HU_{hn}$, i.e., average utility for a healthy person receiving no treatment. We assumed that a person aged 60 would live 22 additional life years, which is equivalent to assuming that these individuals live up to average life expectancy in England. A quality weight of 0.8 was assumed following UK EQ-5D index population norms for similar age groups to those in the target population.[17] To calibrate $HU_{ht}$ it was assumed that the main disutility of a false positive result would derive from the anxiety caused by undertaking a colonoscopy and waiting for results. Following information provided by Public Health England a period of one month could be expected between the moment of a FIT positive result and a colonoscopy result.[16] We further assumed that a patient would lose quality of life during the waiting period according to disutility estimates for moderate anxiety or depression in the UK population.[18] The number of QALYs for untreated sick individuals, $HU_{sn}$, was computed by applying a multiplicative factor to $HU_{hn}$. This factor was taken from the survival ratio of bowel cancer relative to general population survival according to bowel cancer statistics and life tables.[19,20] Finally, the average health utility of those sick receiving treatment, $HU_{st}$, was calibrated by assuming that the diagnostic threshold of 120 µg Hb per $g^{-1}$ faeces, currently used in England, maximises average QALYs in the target population according to the standard approach solution, i.e. maximising expression (1).

**Heterogeneity**

Monte Carlo simulations were used to generate a target population of 1 million subjects (i.e. N=1,000,000). The parameter values that are heterogeneous were generated from the following normal distributions: $p^i \sim N(p, p^{sd})$, $\Delta HU_s^i \sim N(\Delta HU_s, \Delta HU_s^{sd})$ and $\Delta HU_h^i \sim N(\Delta HU_h, \Delta HU_h^{sd})$. Values generated were truncated so that $p^i \in [0, 1]$, $\Delta HU_s^i > 0$ and $\Delta HU_h^i < 0$ while keeping the mean of the distributions unchanged. In the base case analysis, independence between the distributions will be assumed. In this way, the standard deviations were calibrated to predict uptake for England for the FIT assuming that the coefficients of variation are the same for the three distributions.[21] For sake of simplicity, the health utilities under no treatment, $HU_{sn}^i$ and $HU_{hn}^i$, were assumed fixed across the target population and the health utilities under treatment were determined as $HU_{st}^i = HU_{sn}^i + \Delta HU_s^i$ and $HU_{ht}^i = HU_{hn}^i + \Delta HU_h^i$.

## Scenario analyses

Four scenario analyses modify the base case analysis. In Scenario 1, the values of the next parameters are changed separately in three sub-analyses: 1a) The coefficient of variation of $p^i$ , $\Delta HU_s^i$ and $\Delta HU_h^i$; 1b) $\Delta HU_s$, and; 1c) $p$. In this scenario, the same assumptions stated in the proposition of section 2.4 are made, therefore the theoretical results stated are expected to hold. On the contrary, in Scenarios 2 to 4, some of the assumptions are modified on purpose to show specific situations where the theoretical results would change. Scenario 2 modifies the correlation between $p^i$ and $\Delta HU_s^i$, in a multivariate normal distribution model, while keeping $\Delta HU_h^{sd} = 0$. Scenario 3 assumes that uptake behaviour is affected by utility participation as in expression (8). Two sub-analyses are performed: 3a) where a negative participation utility applies to all subjects, and; 3b) where a negative participation utility is applied only to subjects with isoquant slope below the one in expression (2), i.e. $S^i < S$. Finally, Scenario 4 shows the impact of changing the percentage of subjects that are health utility maximisers ($m$), following expression (7), random uptake individuals ($1 - m$) as in expression (9), including two sub-analyses: 4a) where the probability of being health utility maximiser is independent of subjects characteristics, and; 4b) where we assume that the ($1 - m$)% random uptake individuals are the same ($1 - m$)% with the lowest $S^i$ in the target population, i.e. those who obtain the highest expected health utility from undergoing the screening test. In scenario 4, the exogenous participation rate of random uptake individuals (

$u^B$) is also modified to be 0, 0.664 (i.e. the uptake rate predicted for the FIT in the base case analysis), and 1.

## Results

### Base case analysis

Graphical representation of solutions for the maximisation of expression (1) and expression (5) are shown in Figure 4. The two expressions produce different functions to be maximised and distinct optimal solutions (see panels A vs. B). For example, the value of the objective function is greater under the endogenous solution at any point, as only those individuals who regard screening as having a positive expected health utility participate. The endogenous solution is a point on the ROC curve with the pair of TPR and FPR equal to (0.738, 0.479) compared to the standard solution being (0.427, 0.19) as the result of the differences in the isoquant curves (panels C and D). Due to heterogeneity only 50% of the target population would be willing to participate in the screening programme if the endogenous solution is used, 16.4 percentage points lower than the participation associated with the standard approach solution (panels E and F). The base case results are consistent with the proposition stated in section 2.4.

Assuming a target population of 1,000,000 individuals that decide their participation based on their individual expected health utility (expression 7), a screening programme based on the standard solution is estimated to generate a total of 828.1 QALYs compared to a situation with no screening. However, with the endogenous solution 1,002.9 QALYs would be generated compared to no screening, a 21.1% increase in effectiveness.

### Scenario analyses

The results regarding the optimal sensitivity/specificity, level of uptake and QALYs generated for each approach given the uptake behaviour assumed are shown in the supplementary material (Figures A2-A13). As expected, the optimal FPR and TPR under the endogenous approach is always above or equal to the standard solution in the results for Scenario 1 (Figures A2-A4). At the same time, the endogenous solution implies (weakly) lower uptake and higher QALYs generated. The benefit of applying the endogenous approach increases with the level of heterogeneity. For example, the QALYs gained by applying the endogenous

solution will increase from 130 to 676 QALYs if the coefficient of variation was 1 instead of 0.4 (Figure A2). Interestingly, the relationship between QALYs gained and prevalence, or the average treatment effect for sick patients, is non-monotonic (Figures A3-A4). In this sense, the endogenous approach is more valued when the prevalence (or average treatment effect) is not too low or too high. In our example, the endogenous approach will not generate any QALYs if prevalence is close to zero or higher than 20% because both approaches have the same solution in that case (Figure A4).

In scenario 2, we show that the endogenous solution is above the standard approach solution for any value of the Pearson correlation (between the disease risk and the effect of treatment for sick individuals) higher than -0.8 (Figure A5). Only when the distributions of the parameters are very negatively correlated ($corr(\Delta HU_s^i, p^i) \leq -0.8$), the optimal TPR and FPR for the endogenous solution is below the standard approach optimal accuracy. The analysis in Scenario 3a shows that the theoretical result stated in the Proposition of section 2.4 holds even when we account for a fixed negative participation utility across the target population (Figure A6). In that case, uptake is not a negative function of FPR/TPR, in contrast to the case when uptake is completely driven by expected health utility. When the participation utility is equal or below -0.008 and affects only subjects with the lowest isoquant slopes in Scenario 3b, the optimal accuracy of the test for the endogenous solution entails a lower TPR and FPR than for the standard approach solution (Figure A7). Interestingly, in scenario 3b, the optimal accuracy under the endogenous solution sharply declines for values below -0.008. The results of Scenario 4 show that the value, in terms of QALY gain, of using the endogenous solution, tends to decrease with the proportion of random uptake subjects in the population (Figures A8-A13). Nonetheless, the endogenous solution is always equal or above the standard approach optimal accuracy when the probability of being health utility maximiser is independent of the individual's characteristics (Scenario 4a) (Figures A8-A10). However, when only the individuals with the lowest health utility isoquant choose randomly (Scenario 4b), and the exogenous uptake rate ($u^B$) is 0, the endogenous solution could be below the standard approach one if $m$ is sufficiently low (Figure A11).

# Discussion and conclusion

## What this study has found

The present paper illustrates the economic problem of valuing a screening test and choosing the optimal sensitivity and specificity when uptake is considered endogenous. If individuals are health utility maximisers, the choice of the accuracy of a test along the ROC curve will affect their participation. We have shown that choosing a diagnostic threshold, with its corresponding TPR and FPR, and valuing a screening test using the standard approach solution can be suboptimal if uptake is endogenous. The use of an endogenous approach can lead to a more optimal situation particularly when heterogeneity is high and the average utility of receiving treatment is moderate. The intuition behind this result is that in those situations there might be some patients that receive only a small benefit from treatment and their optimal decision may be not participating, so it becomes more important to integrate individuals' uptake decisions into the economic evaluation of the screening programme. Also, the endogenous uptake approach is most valuable in cases where the prevalence of the disease is very high, where the best solution is to send all individuals for treatment or further investigation (TPR=1, FPR=1), or very low, where the best solution is to send no-one for treatment or further investigation (TPR=0, FPR=0).

We have shown that the optimal combination of TPR and FPR for a screening test is not the same one that maximises uptake. In other words, the decision maker should not change the diagnostic threshold for a test, for example the amount of occult blood in the FIT that will lead to a decision of further investigation, to increase uptake if the objective is to maximise the average utility for the target population. We may wrongly think that we should choose a threshold that maximises participation so that more patients benefit from early detection of a disease. However, this would be at the expense of reducing the utility of those individuals who value the test the most, which is why an uptake-maximising strategy may not be optimal. Notice that this is not stating that public health bodies should not aim at improving uptake of those individuals that would benefit from screening, given the level of TPR and

FPR, but decide not to participate because of other barriers. For example, imagine an individual who has decided to not participate in a test that is too time-consuming. Any public health intervention that provided an incentive for the person's participation or an improvement in the administration of the test could be consistent with a general objective of maximising the average health utility of the target population.

Another result from the present analysis is that, under a wide range of assumptions in the practical example, the optimal combination of TPR and FPR under an endogenous uptake analysis is at a higher point on the ROC curve than the optimal solution under the standard approach. This means that both the TPR and FPR are higher. If individuals decide their participation based on the accuracy of the test (endogenous uptake), we should reduce the diagnostic threshold, compared to the standard solution, in order to increase the number of true positive even if this is at the expense of increased false positive cases. This is an intuitive result that comes from the fact that, under endogenous uptake, those who participate are individuals who value treatment the most (more than the average person in the target population) and therefore will improve their health utility (on average) if we move to a higher TPR point in the ROC curve. The standard approach is not able to account for changes in uptake, failing to adapt to those individuals who actually participate in the screening programme.

## Strengths

Our analysis is based on both theoretical and numerical methods to show the importance of considering endogenous uptake when designing screening programmes. On the one hand, the theoretical framework illustrates how heterogeneity can be incorporated within economic evaluation analysis by making use of individual decision-making models such as expected utility theory. This heterogeneity can be explicitly set as a variable in the broad economic value function of a screening programme where the benefits and costs for the whole population of interest is included. On the other hand, the example of bowel cancer screening in England is based on plausible data of the performance of the FIT and the national screening outputs in terms of diagnostic results of participants and uptake rates. Interestingly, in the base case analysis, i.e., using calibrated parameters derived from real-world data, the number of QALYs obtained by the bowel screening programme could increase by 21.1% if

heterogeneity is considered. If we do not consider heterogeneity and uptake behaviour when valuing screening tests, we may systematically under-value a screening programme.

This paper contributes to a broad set of studies in health economics that analyse the importance of patient heterogeneity in economic evaluations.[22-26] Our analysis is particularly related to Kim and Basu (2017) where the cost effectiveness at the population level is measured for a treatment that would be adopted differently by heterogeneous patient subgroups.[27] Our study can be considered within this framework where the treatment to be evaluated is a screening test provided to a target population and uptake behaviour is specified as endogenous to the test accuracy. In this sense, our study is also in line with the causal inference literature that considers self-selection into and out of treatment, where the treated people can experience systematically higher or lower treatment effects than the average effect across the whole population.[28,29]

## Limitations

The results obtained are restricted to screening programmes where the same unique diagnostic threshold (i.e. the same optimal TPR and FPR) is going to be used for the whole target population, as in the BCSP in the UK. Therefore, our analysis would not apply to the case where the diagnostic threshold is optimally set for each patient individually. Notice however that this alternative scenario has its own limitations in terms of costs for health systems and patients. For example, it would require higher interaction between doctors and patients to discuss treatment options individually. It would also require a high cognitive effort made by patients in understanding ROC curves, since they would need to know all the alternatives in the choice set before they decided to participate in the screening.

Future research could rethink the assumptions and limitations of this work so that a debate could take place. For example, the present analysis assumes the decision maker (public health body) is primarily interested in maximising the average health utility, however a full economic evaluation should consider costs for the healthcare system as well. In this sense, the main interesting feature of a full economic evaluation, compared to the present analysis, would be the asymmetry between health authorities and individuals about relevant outcomes. The decision maker (public health authorities) would care about both health outcomes and healthcare costs in their objective functions; however, individuals will only consider their

health utilities or out of pocket costs in their utility functions. In addition, the data we use for our real-life application to the BCSP in the UK is based on derived parameter estimates and some specific assumptions, so this analysis should not be relied upon for decision making about that programme.

Finally, the endogenous solution is affected by the assumptions made about uptake behaviour. For example, uptake behaviour may be affected by non-health-utility factors.[10,11] We have tried to relax the expected health utility assumption by including the concept of participation utility, in Scenario 3, and random uptake, in Scenario 4; but there may well be alternative specifications of uptake decisions that could be more realistic. First, subjects could be wrong about their risk of being sick: they may overestimate or underestimate the probability of having a specific disease. Furthermore, imagine if those who overestimate the risk are those who really are less likely to develop the disease. For example, people who are under-informed may be those who face a high risk of disease. In this sense, there is evidence that socioeconomic status is correlated with cancer screening uptake which at the same time is correlated with actual risk of cancer.[30,31] Nevertheless, some people may have the ability to identify, at least to some extent, their level of risk as suggested by Palmer et al.'s study (2014), where "feeling well was associated with low perceived relevance of screening" in colorectal cancer.[10] Also, the way people value probabilities may not be linear but rather S-shaped as behavioural economics suggests.[32,33] For example, individuals could value much more the reduction of probability of a false positive from 5% to 0% than the same reduction from 10% to 5%.[34]

## Conclusion

When uptake is endogenous to the accuracy of the screening test, individual heterogeneity and uptake decisions are relevant factors to consider in the problem of finding an optimal combination of sensitivity and specificity; this approach could improve the value of screening programmes.

**Ethical considerations**

n/a

**Consent to participate**

n/a

**Consent for publication**

n/a

**Declaration of conflicting interests**

None

**Funding statement**

None

**Data availability**

n/a

REFERENCES

1. Terris-Prestholt F, Quaife M, Vickerman P. Parameterising user uptake in economic evaluations: the role of discrete choice experiments. *Health economics*. 2016;25:116-123.
2. Laking G, Lord J, Fischer A. The economics of diagnosis. *Health economics*. 2006;15(10):1109-1120.
3. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Medical Decision Making*. 1988;8(4):279-289.
4. Groothuis-Oudshoorn CG, Fermont JM, van Til JA, IJzerman MJ. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. *BMC medical informatics and decision making*. 2014;14(1):18.
5. Von Wagner C, Baio G, Raine R, et al. Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England. *International journal of epidemiology*. 2011;40(3):712-718.
6. NHS-digital. Breast Screening Programme, England - 2016-17. England2018.
7. NHS-England. Cervical Screening Programme, England - 2016-17. United Kingdom2017.
8. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. *New England Journal of Medicine*. 1975;293(5):211-215.
9. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. Oct 1978;8(4):283-98. doi:10.1016/s0001-2998(78)80014-2
10. Palmer C, Thomas M, Von Wagner C, Raine R. Reasons for non-uptake and subsequent participation in the NHS Bowel Cancer Screening Programme: a qualitative study. *British journal of cancer*. 2014;110(7):1705.
11. Chapple A, Ziebland S, Hewitson P, McPherson A. What affects the uptake of screening for bowel cancer using a faecal occult blood test (FOBt): a qualitative study. *Social science & medicine*. 2008;66(12):2425-2435.
12. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *New England Journal of Medicine*. 1980;302(20):1109-1117.
13. Cooper JA, Parsons N, Stinton C, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. *British journal of cancer*. 2018;118(2):285.
14. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Medicine; 2003.
15. National Health Service. NHS bowel cancer screening: helping you decide (FIT kit). Public Health England. Accessed 22/02/2021, 2021. https://www.gov.uk/government/publications/bowel-cancer-screening-benefits-and-risks/nhs-bowel-cancer-screening-helping-you-decide
16. National Health Service. Bowel cancer screening: having a colonoscopy (FIT). Public Health England. Accessed 22/02/2021, 2021. https://www.gov.uk/government/publications/bowel-cancer-screening-colonoscopy/bowel-cancer-screening-having-a-colonoscopy-fit
17. Szende A, Janssen B, Cabases J. *Self-reported population health: an international perspective based on EQ-5D*. Springer Netherlands Dordrecht; 2014.
18. Dolan P. Modeling valuations for EuroQol health states. *Medical care*. 1997:1095-1108.
19. Office for National Statistics. National life tables: England. Office for National Statistics. Accessed 22/02/2021, https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandreferencetables/current
20. Cancer Research UK. Bowel cancer survival statistics. Accessed 22/02/2021, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#ref-

21. Moss S, Mathews C, Day TJ, et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. *Gut*. 2017;66(9):1631-1644. doi:10.1136/gutjnl-2015-310691
22. Phelps CE. Good technologies gone bad: how and why the cost-effectiveness of a medical intervention changes for different populations. *Med Decis Making*. 1997;17(1):107-17. doi:10.1177/0272989X9701700113
23. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Medical decision making*. 2002;22(3):218-227.
24. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Economics*. 2003;12(5):421-427.
25. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Medical Decision Making*. 2007;27(2):112-127.
26. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. *Medical Decision Making*. 2014;34(8):951-964.
27. Kim DD, Basu A. New metrics for economic evaluation in the presence of heterogeneity: focusing on evaluating policy alternatives rather than treatment alternatives. *Medical Decision Making*. 2017;37(8):930-941.
28. Basu A, Heckman JJ, Navarro-Lozano S, Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health economics*. 2007;16(11):1133-1157.
29. Mogstad M, Santos A, Torgovitsky A. Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*. 2018;86(5):1589-1619.
30. Bang JY, Yadegarfar G, Soljak M, Majeed A. Primary care factors associated with cervical screening coverage in England. *Journal of Public Health*. 2012;34(4):532-538.
31. Maheswaran R, Pearson T, Jordan H, Black D. Socioeconomic deprivation, travel distance, location of service, and uptake of breast cancer screening in North Derbyshire, UK. *Journal of Epidemiology & Community Health*. 2006;60(3):208-212.
32. Tversky A, Kahneman D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*. 1992;5(4):297-323.
33. Kahneman D. Prospect theory: An analysis of decisions under risk. *Econometrica*. 1979;47:278.
34. Bleichrodt H, Pinto JL. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management science*. 2000;46(11):1485-1496.

**Table 1.** Parameters in the base case analysis.

| Parameter | Sick state | Healthy state |
|---|---|---|
| ROC curve (FIT) | $\theta_s \sim N(96.9, 63)$ | $\theta_h \sim N(53.3, 63)$ |
| QALYs under no treatment | $HU_{sn}^i = HU_{sn} = 11.008$ | $HU_{hn}^i = HU_{hn} = 17.6$ |
| QALYs under treatment | $HU_{st}^i = HU_{sn}^i + \Delta HU_s^i$ | $HU_{ht}^i = HU_{hn}^i + \Delta HU_h^i$ |
| | $\Delta HU_s^i \sim N(0.1425, 0.0026)$ | $\Delta HU_h^i \sim N(-0.0060, 0.0617)$ |
| Prevalence | $p^i \sim N(0.0285, 0.1233)$ | |

Note. Figures shown have been rounded to four decimal places. N(mean, sd) stands for normal distribution with the given mean and standard deviation. Standard deviations of $p^i$, $\Delta HU_s^i$ and $\Delta HU_h^i$ are calibrated so that the coefficient of variation is 0.443.

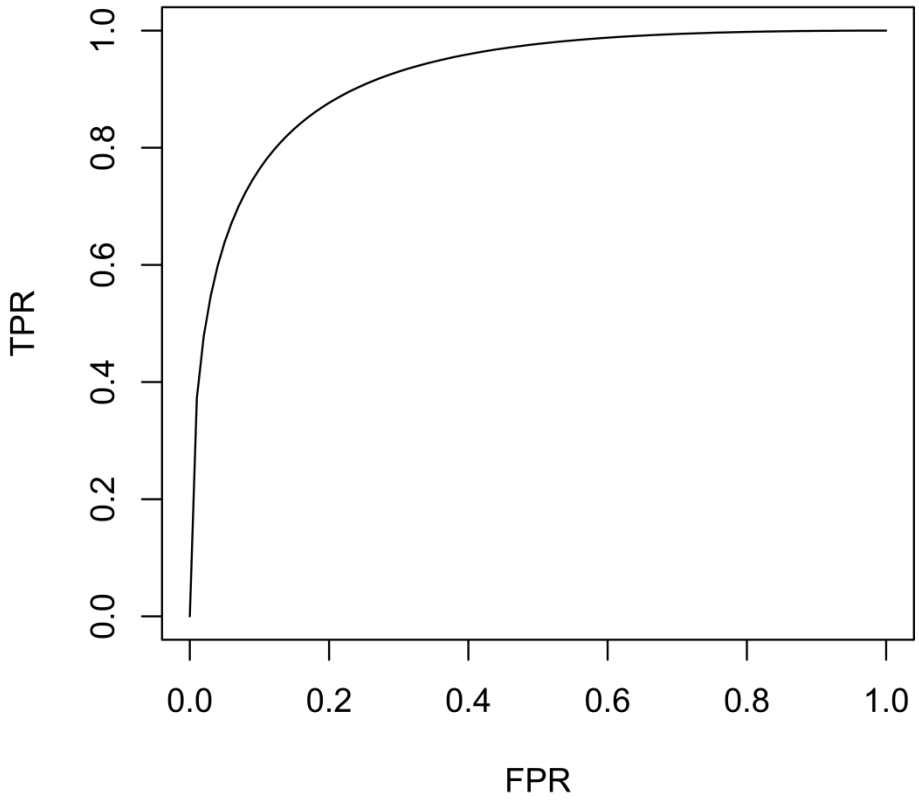


Figure 1/ROC curve. FPR – false positive rate; TPR – true positive rate

50x50mm (3000 x 3000 DPI)

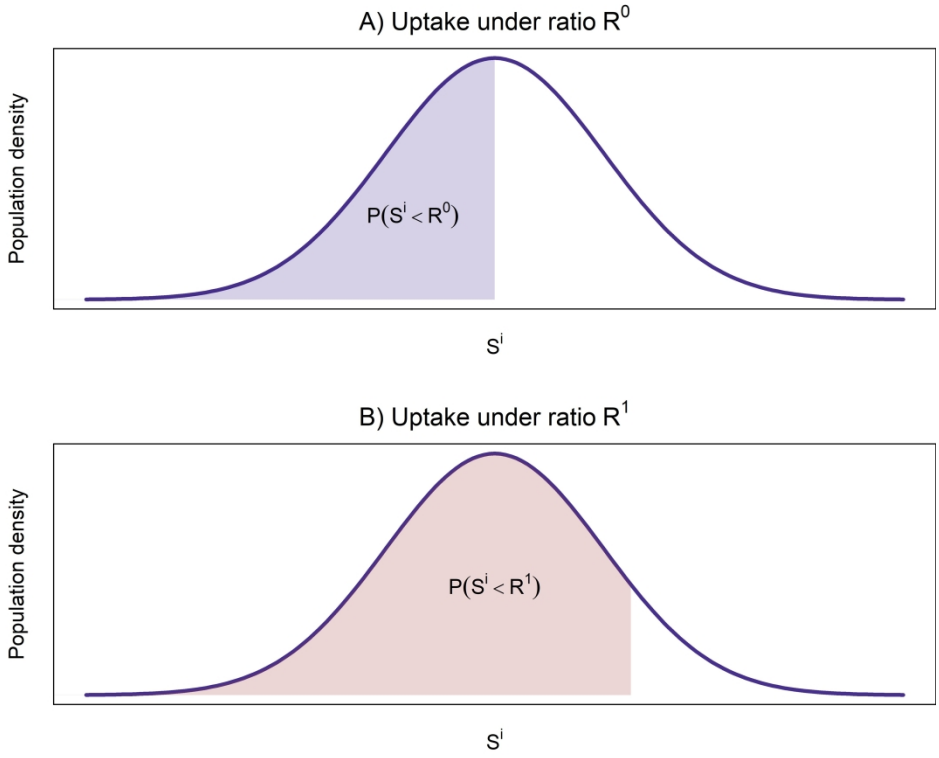


Figure 2/Screening uptake under a low and high ratio of TPR and FPR (R0<R1)

159x126mm (750 x 750 DPI)

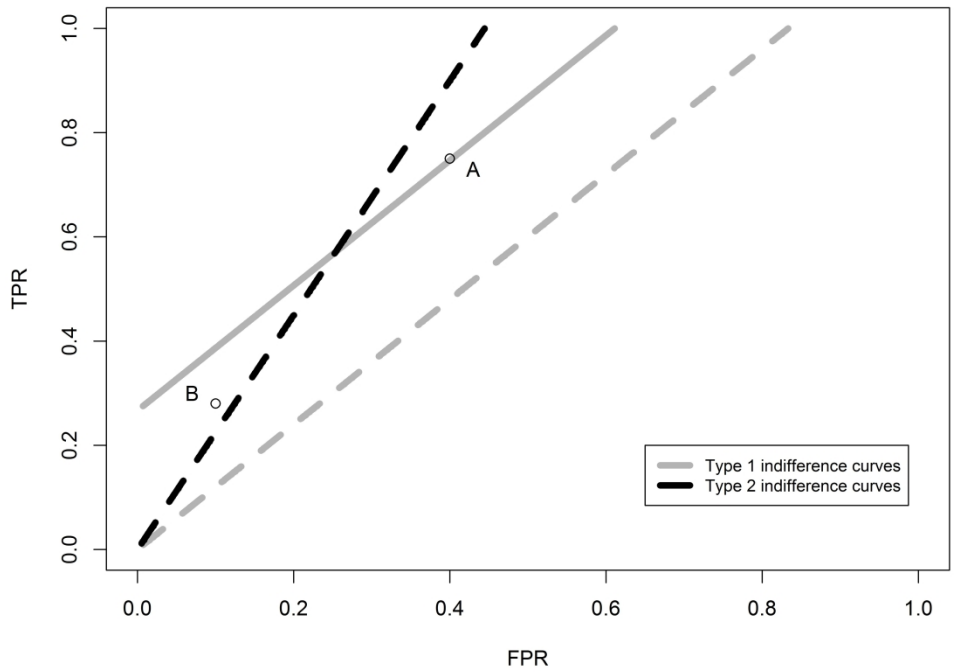


Figure 3/Choice between A and B. Indifference curves for individual type 1 (grey) and individual type 2 (black). TPR – true positive rate; FPR – false positive rate

159x126mm (750 x 750 DPI)

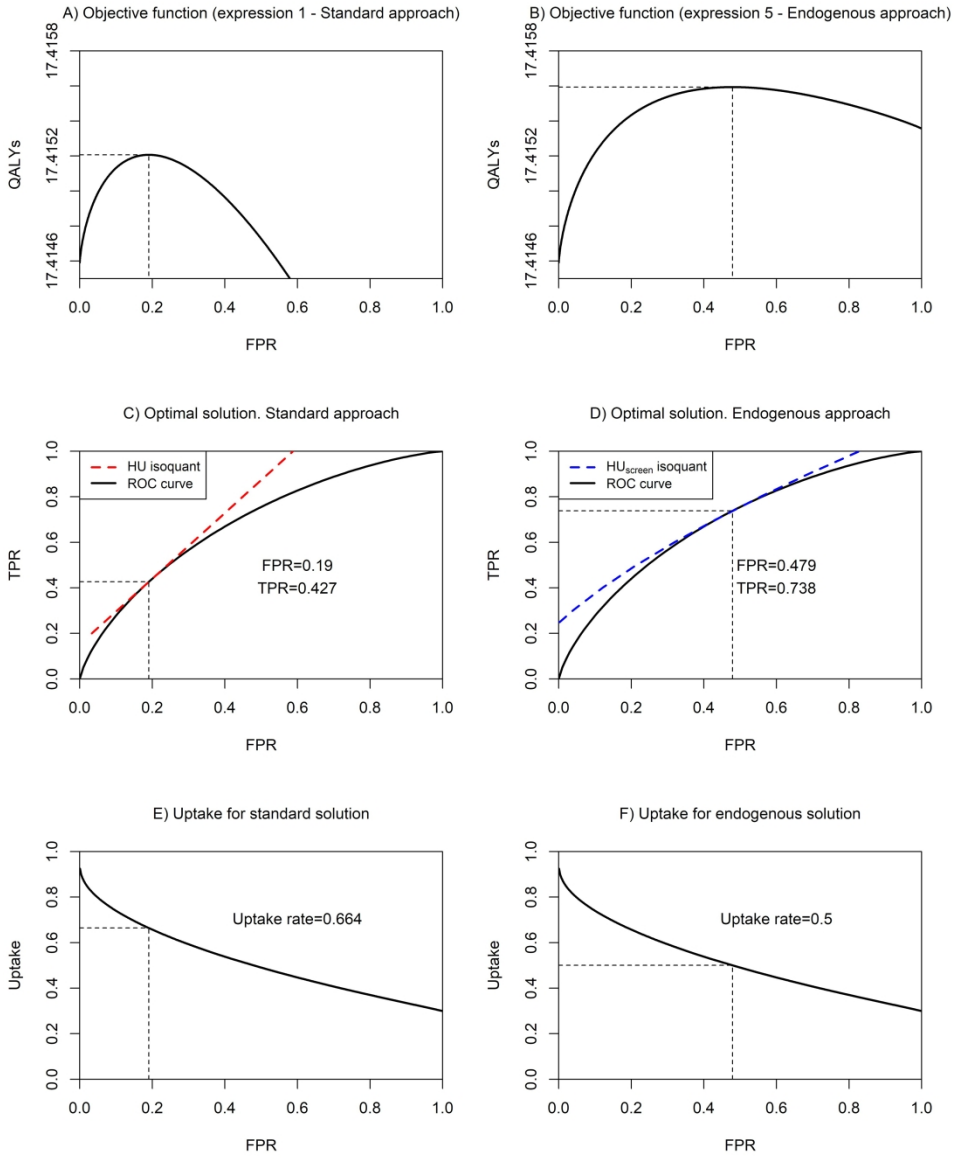


Figure 4/Representation of the optimal solutions under standard approach and endogenous uptake approach. FPR – false positive rate; TPR – true positive rate; ROC – receiver operating characteristic

159x198mm (630 x 630 DPI)

# Levelling up: treating uptake as endogenous may increase the value of screening programmes

Supplementary material

Supplementary material

## A1. Calibration of parameters for the bowel screening example in England

FIT ROC curve

The binormal model was calibrated to fit the empirical ROC curve estimated in previous work by Cooper et al. (2018). [1] The mean occult blood for the healthy and sick population was obtained from the median reported in Figure S1 in their supplementary analysis. [1] Standard deviation of the biomarker is chosen to minimise the sum of the squared errors between the fitted and empirical curve. The fitted ROC curve (Figure A1) is given by the binormal model where the distribution of the diagnostic variable (occult blood) was fitted to $\theta_s \sim N(96.9, 63)$ and $\theta_h \sim N(53.3, 63)$, for the sick and healthy population.

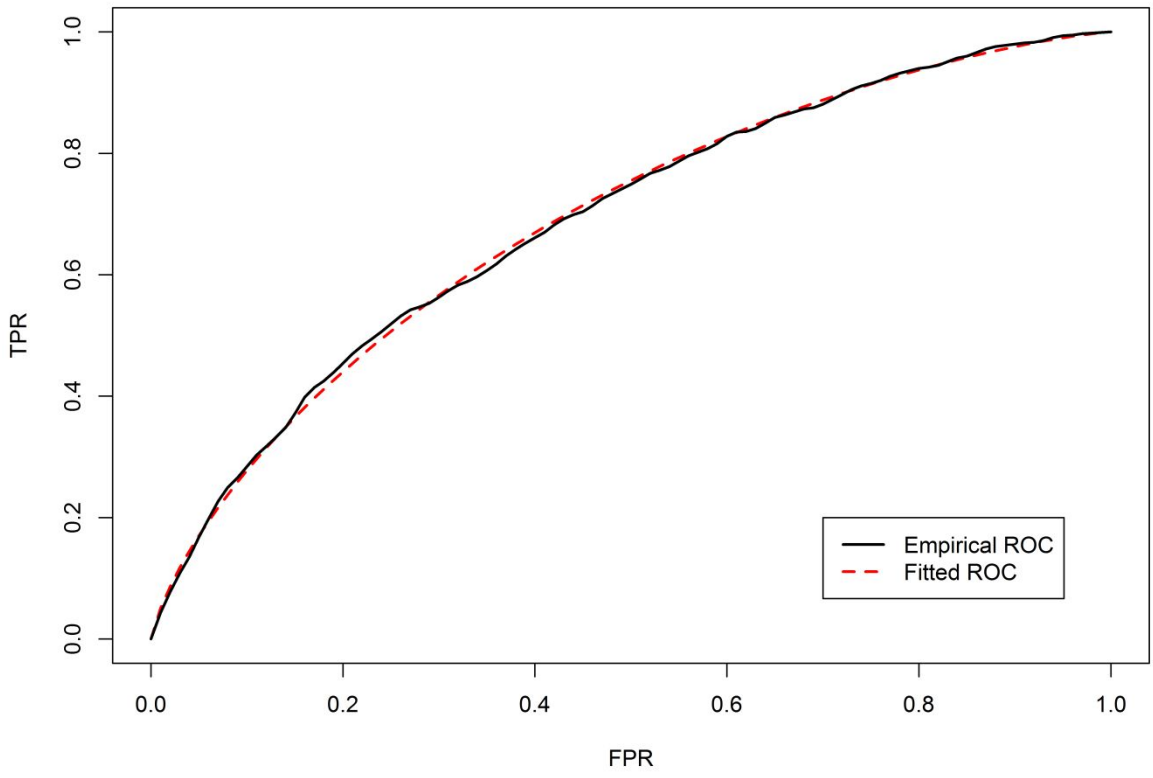


*Figure A1. Digitised empirical ROC curve (Cooper et al., 2018) and binormal model fitted ROC curve for the FIT.*

Prevalence

The formula used to compute prevalence of bowel cancer or advanced adenoma is:

$$p = \frac{P(TP)}{TPR}, \quad \text{(A1)}$$

i.e. the ratio of the probability of detecting a true positive case P(TP) in the screened population and the true positive rate.

Following UK Government reports on bowel cancer screening in England we assumed that 2% of screened population are tested positive, with 62% of them being true positive.[i] So, the probability of being a true positive in the screened population is 1.24%. Given that the TPR at the diagnostic threshold used (120µg/g) is 43.53% we can compute the prevalence as $p = \frac{0.0124}{0.4353} = 0.028484$.

Health utilities

The value of $HU_{hn}$ is computed as the multiplication of 22 years of assumed survival years and 0.8 of assumed quality weight, $HU_{hn} = 22 \times 0.8 = 17.6$. A disutility for a false positive result is assumed to be similar to the disutility estimates for moderate anxiety or depression estimated for the UK population, 0.071 as estimated by Dolan (1997).[2] Since we assume that this disutility lasts for one month, we can compute the utility for false positive cases as $HU_{ht} = 17.6 - \left(\frac{31}{365}\right) \times 0.071 = 17.594$. The number of QALYs for untreated sick patients, $HU_{sn}$, was computed by applying a multiplicative factor to $HU_{hn}$ as $HU_{sn} = 17.6 \times 0.63 = 11.008$, where 0.63 is the ratio of the 10-year survival of bowel cancer patients relative to 60-year-old person in the general population according to bowel cancer statistics and life tables.[ii] Finally,

[i] See:
- https://www.gov.uk/government/publications/bowel-cancer-screening-benefits-and-risks/nhs-bowel-cancer-screening-helping-you-decide
- https://www.gov.uk/government/publications/bowel-cancer-screening-colonoscopy/bowel-cancer-screening-having-a-colonoscopy-fit

[ii] See:
- https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandreferencetables/current
- https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#ref-

$HU_{st}$ is set to 11.23 in such a way that if we assume the rest of the parameters as in Table 1 then the diagnostic threshold used in England, 120 µg Hb per $g^{-1}$ faeces, provides the optimal combination of TPR=0.427 and FPR=0.19, along the fitted ROC curve, that maximises expression (1), i.e. we assume that the diagnostic threshold used is the optimal solution under the standard approach to economics of diagnostics.

Heterogeneity of preferences

Heterogeneity parameters are calibrated so that our model matches estimated uptake for the FIT. Moss et al. (2017) estimate an uptake of 66.4% of the target population.[3] Our model needs to replicate this uptake as the probability of participating in our heterogenous population, in such a way that $P\left(S^i = \frac{-(1-p^i)(\Delta HU_h^i)}{p^i(\Delta HU_s^i)} < R = \frac{TPR}{FPR}\right) = 0.664$, or equivalently $\frac{\sum_{i=1}^{N} u^i}{N} = 0.664$. Under the base case analysis, this participation is achieved when the coefficient of variation is 43.3% for the three distributions: $p^i \sim N(p, p^{sd})$, $\Delta HU_s^i \sim N(\Delta HU_s, \Delta HU_s^{sd})$ and $\Delta HU_h^i \sim N(\Delta HU_h, \Delta HU_h^{sd})$. Consequently, the standard deviations are computed as: $p^{sd} = 0.433 \times p = 0.12334$, $\Delta HU_s^{sd} = 0.433 \times \Delta HU_s = 0.0026110$ and $\Delta HU_h^{sd} = 0.433 \times \Delta HU_h = 0.06168951$. For sake of simplicity, the health utilities under no treatment, $HU_{sn}^i$ and $HU_{hn}^i$, were assumed fixed across the target population. Notice that the main result in Section 2.4 of the main manuscript is valid when heterogeneity is present in any of the health utility parameters, so this assumption is enough to show the expected differences between the standard and endogenous solution.

The parameters values that are heterogeneous were generated from the following normal distributions: $p^i \sim N(p, p^{sd})$, $\Delta HU_s^i \sim N(\Delta HU_s, \Delta HU_s^{sd})$ and $\Delta HU_h^i \sim N(\Delta HU_h, \Delta HU_h^{sd})$. Values generated were truncated so that $p^i \in [0, 1]$, $\Delta HU_s^i > 0$ and $\Delta HU_h^i < 0$ while keeping the mean of the distributions unchanged. For example, if for a subject $i$ the replicated value of disease risk was $p^i < 0$ it was replaced by 0. If this happened, truncation was also applied to subjects with the highest values in the distribution of $p^i$ so that the same percentage of truncation was applied from below and from above. Particularly, the highest values of $p^i$ were replaced by the value of the $(1 - p_0)^{th}$ percentile in the distribution, where $p_0$ represents the percentile of 0. The same method of truncation was applied if $p^i > 0$, $\Delta HU_s^i > 0$ and $\Delta HU_h^i < 0$. This procedure warrants preservation of the mean and the symmetry of the

distribution. In the base case analysis, the Monte Carlo replications had 1.19% of subjects with $p^i < 0$, 0% of subjects with $p^i > 1$, 1.19% of subjects with $\Delta HU_s^i < 0$, and 1.2% of individuals with $\Delta HU_h^i < 0$.

## A2. Scenario analyses

Scenario 1a

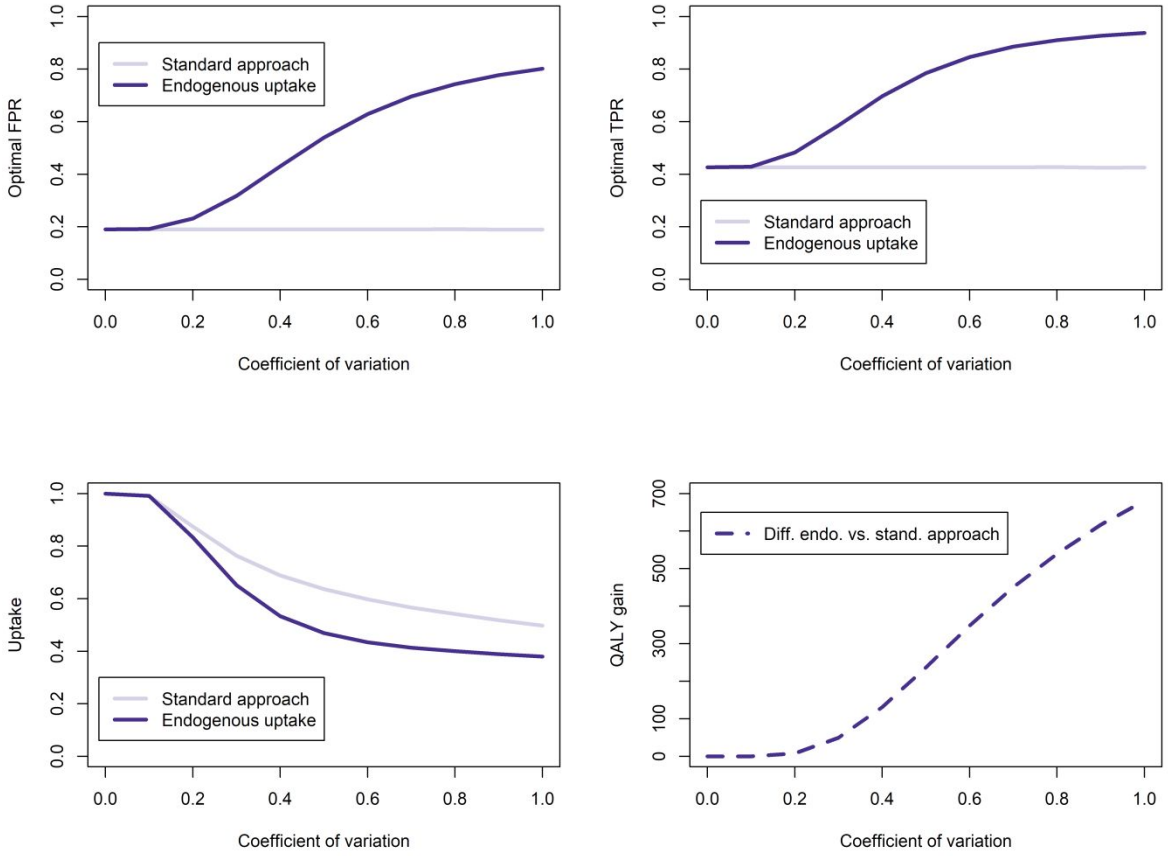


Figure A2. Comparison of the two approaches under different values of the coefficient of variation of $p^i$, $\Delta HU_s^i$ and $\Delta HU_h^i$. Assuming a target population of 1,000,000 subjects. (Coefficient of variation changed in steps of 0.1)

Scenario 1b

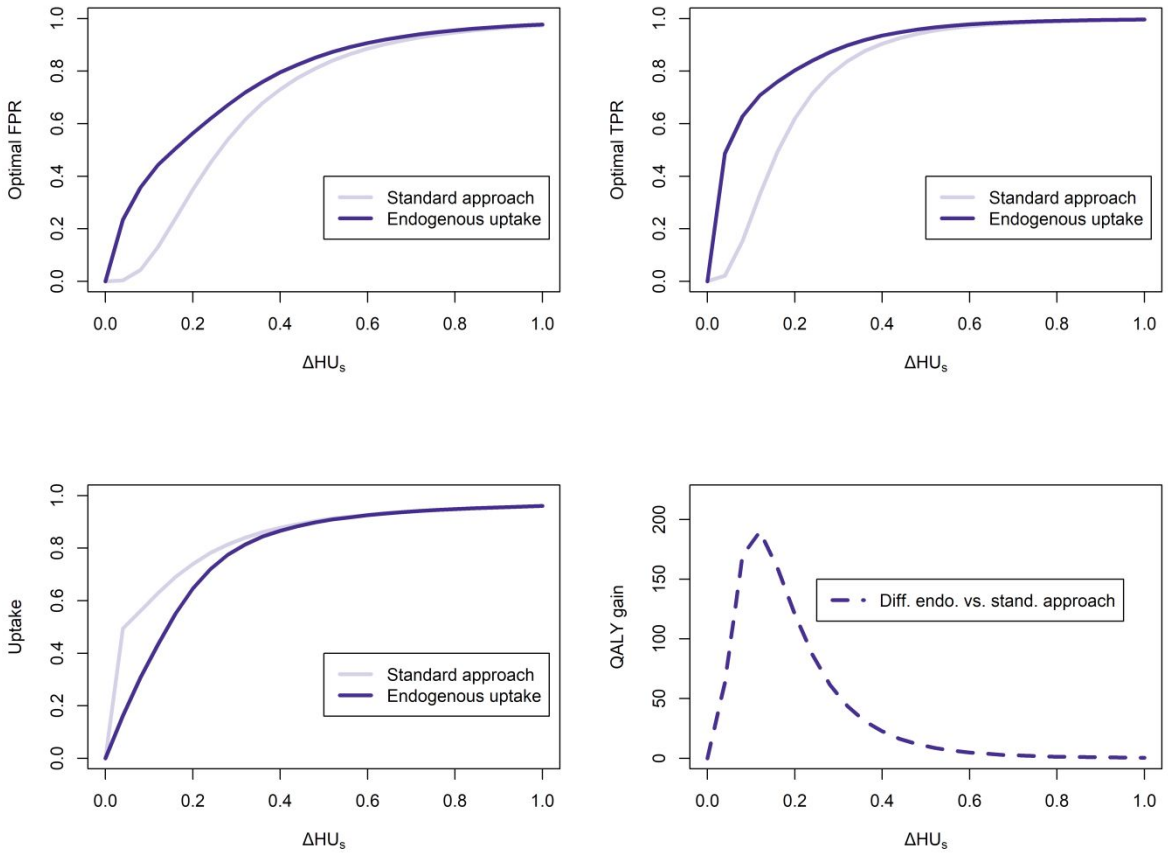


Figure A3. Comparison of the two approaches under different values of $\Delta HU_s$. Assuming a target population of 1,000,000 subjects. ($\Delta HU_s$ changed in steps of 0.04)

Scenario 1c

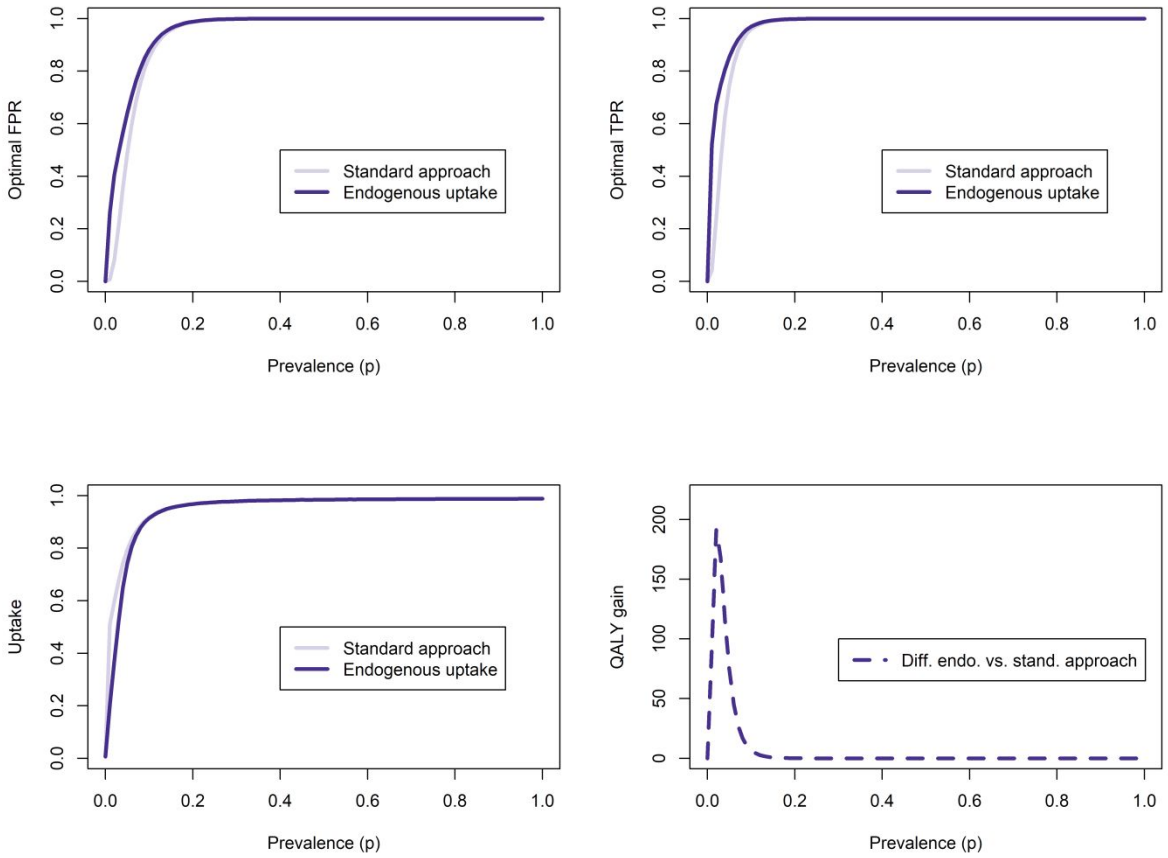


Figure A4. Comparison of the two approaches under different values of prevalence (p). Assuming a target population of 1,000,000 subjects. ($p$ changed in steps of 0.01)

Scenario 2

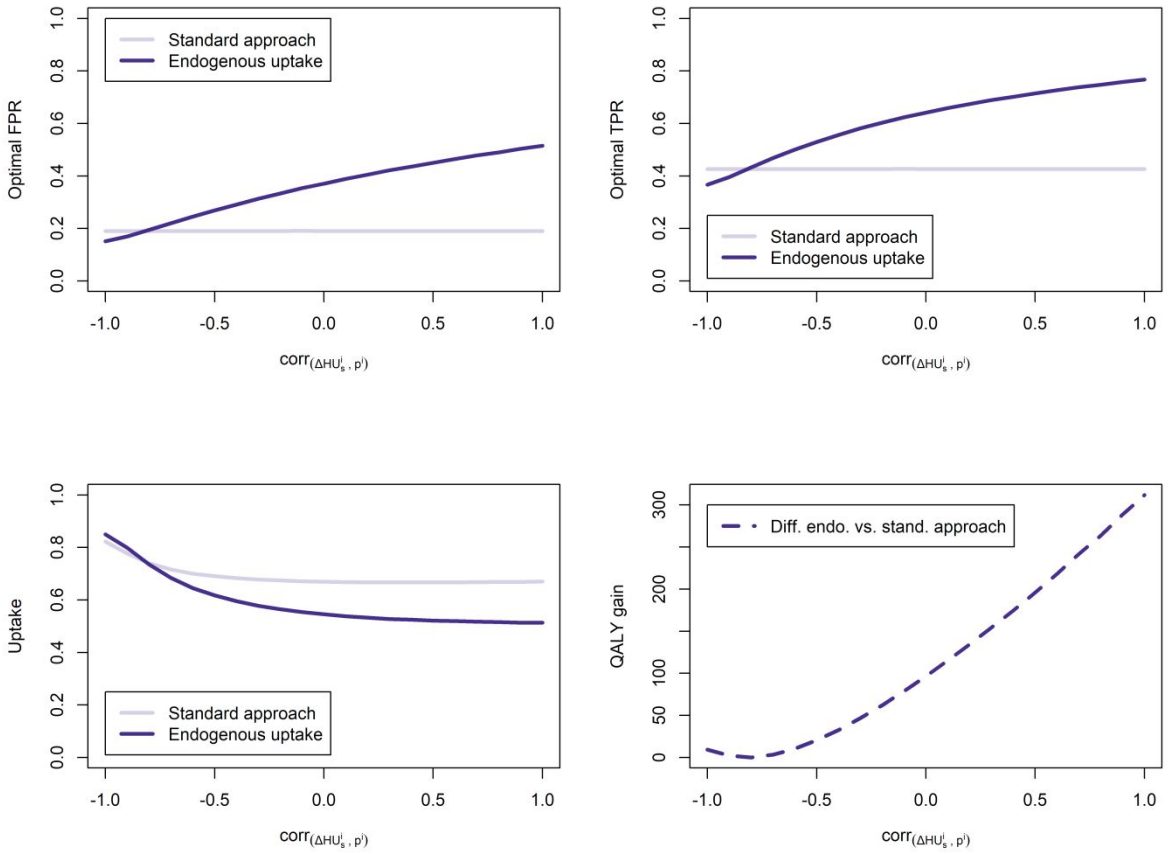


Figure A5. Comparison of the two approaches under different values of the Pearson correlation coefficient between $\Delta HU_s^i$ and $p^i$. Assuming a target population of 1,000,000 subjects. In this scenario $\Delta HU_h^{sd} = 0$. ($corr_{(\Delta HU_s^i, p)}$ changed in steps of 0.1)

Scenario 3a

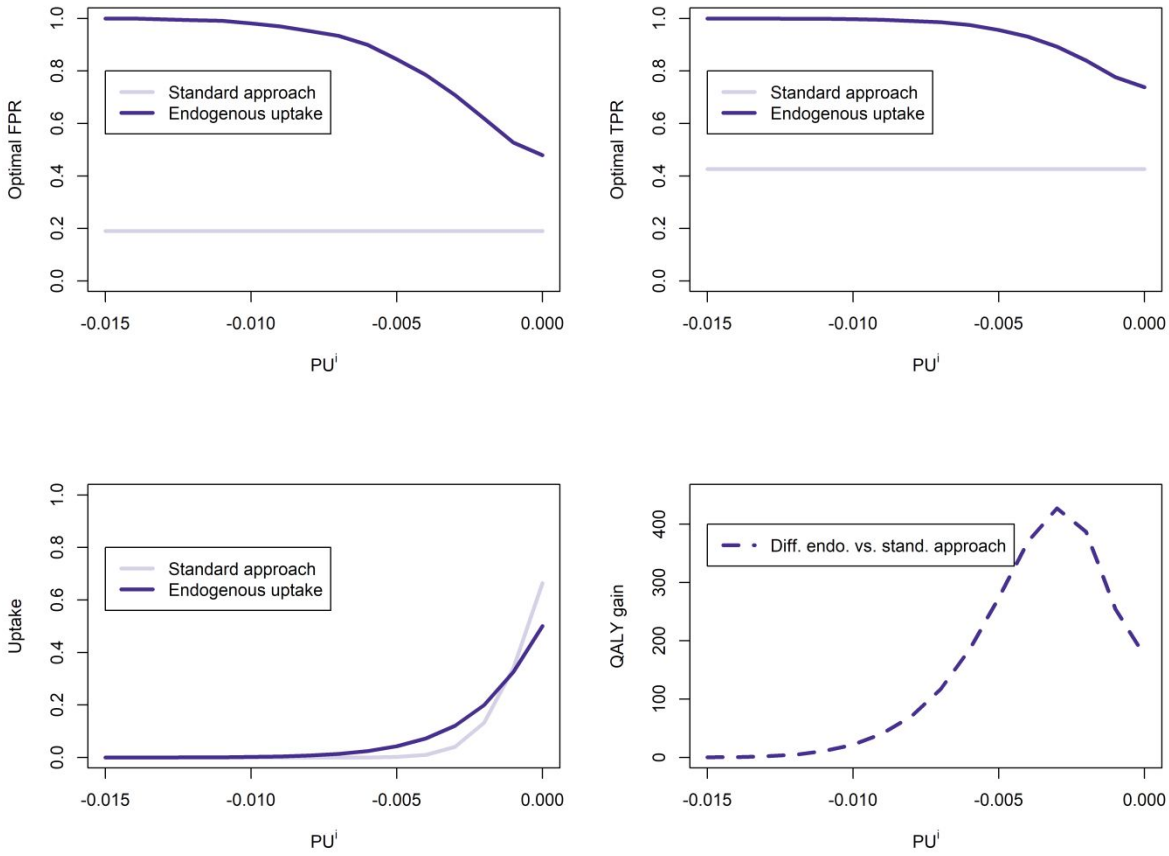


Figure A6. Comparison of the two approaches under different values of participation utility (equal participation utility for all subjects). Uptake behaviour given by expression (8). Assuming a target population of 1,000,000 subjects. ($PU^i$ changed in steps of 0.001)

Scenario 3b

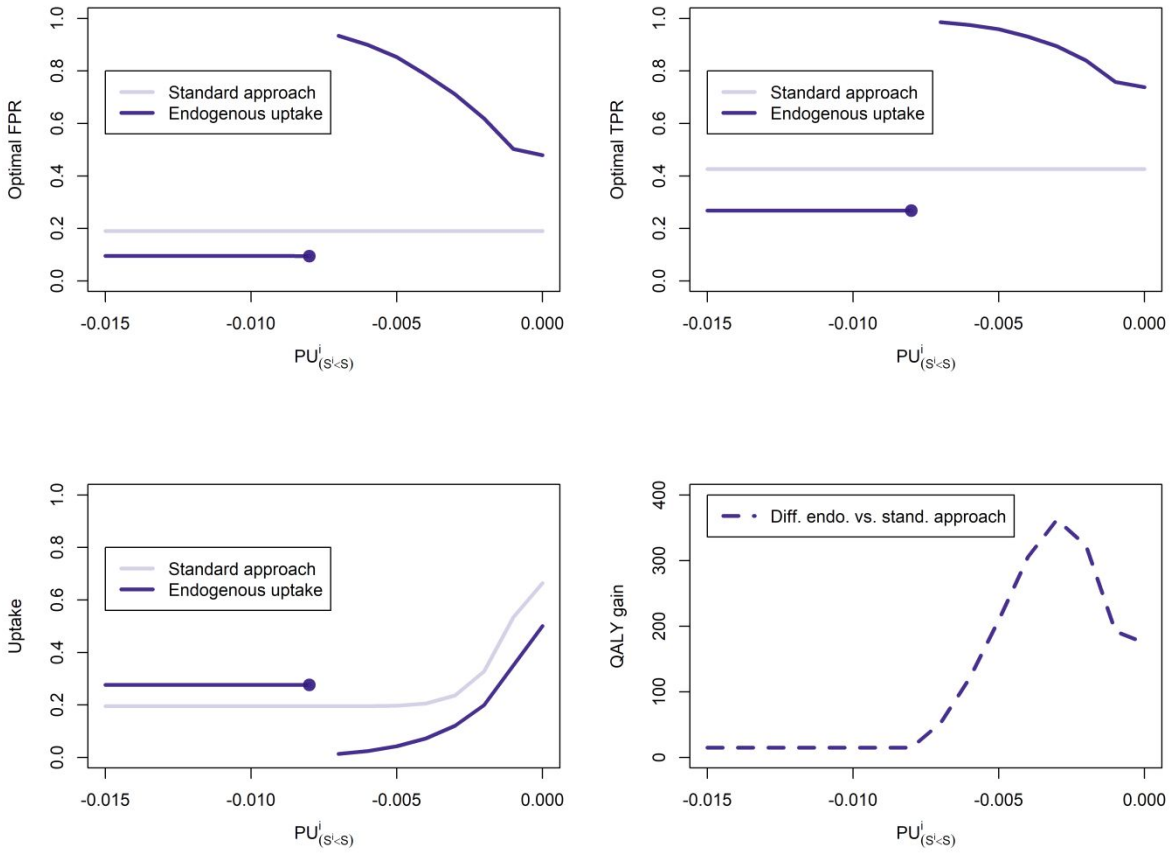


Figure A7. Comparison of the two approaches under different values of participation utility. Participation utility applied only to subjects with $S^i < S$. Uptake behaviour given by expression (8). Assuming a target population of 1,000,000 subjects. ($PU^i_{(S^i<S)}$ changed in steps of 0.001)

Scenario 4a

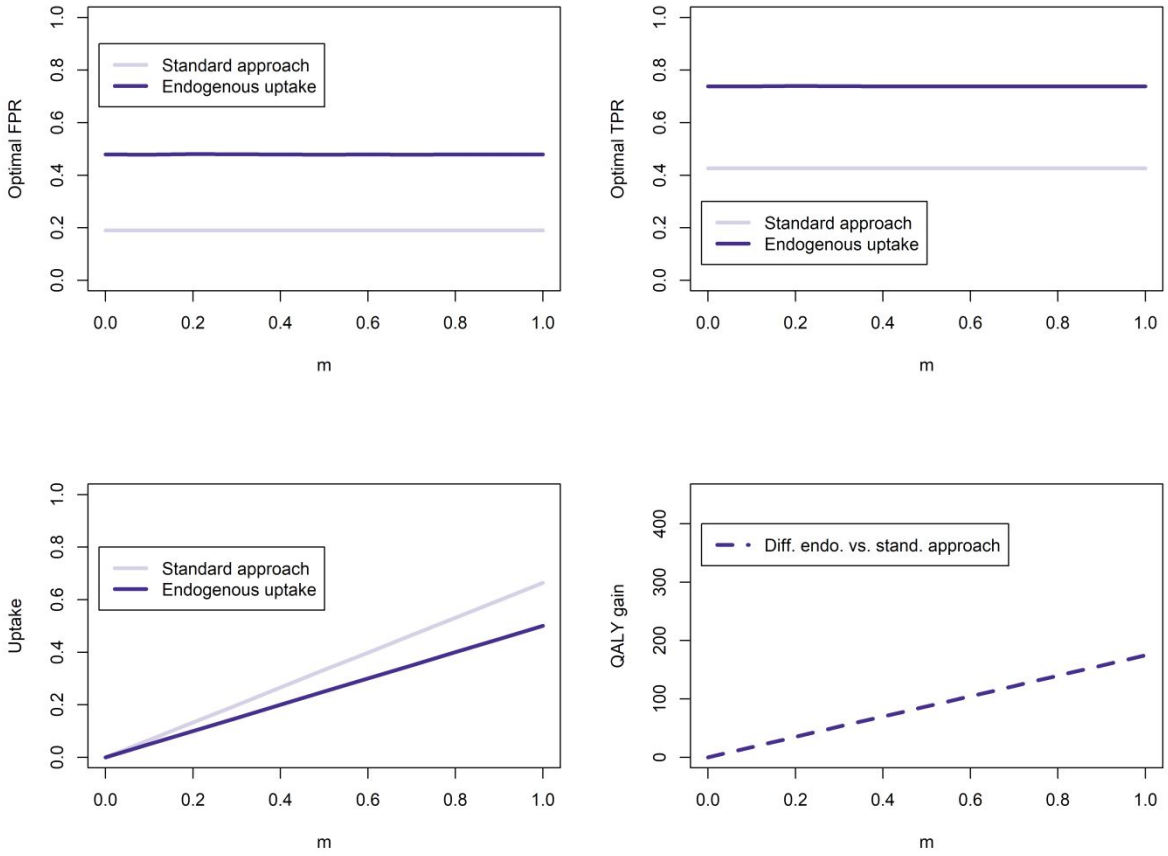


Figure A8. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The probability of being health utility maximiser is independent of subject's characteristics. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 0. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

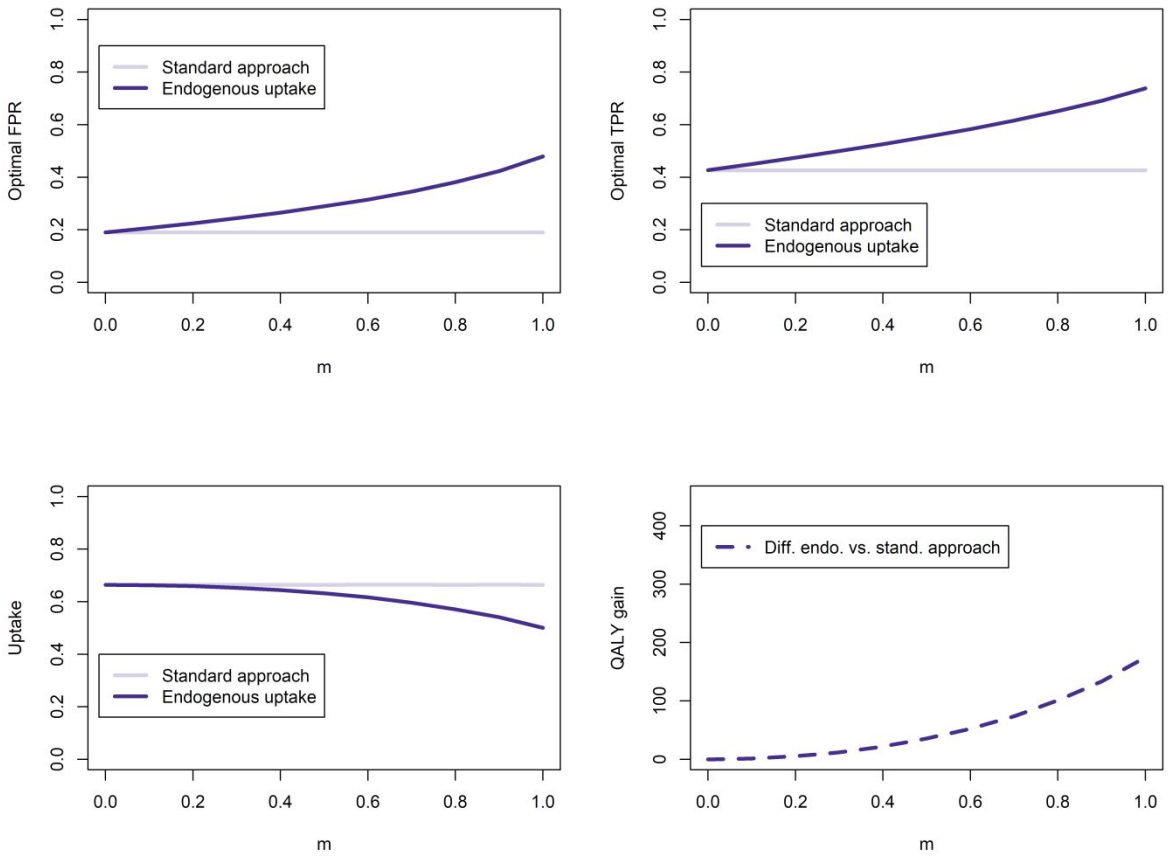


Figure A9. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The probability of being health utility maximiser is independent of subject's characteristics. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 0.664. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

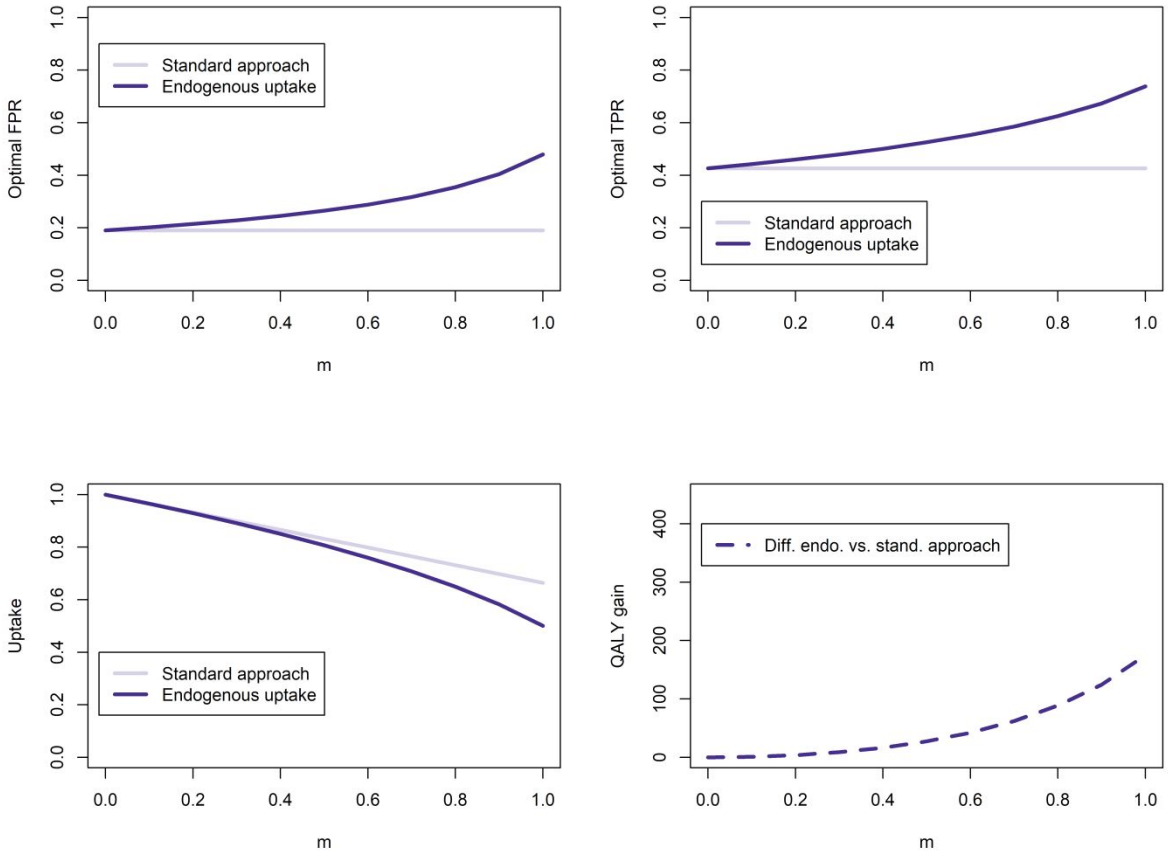


Figure A10. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The probability of being health utility maximiser is independent of subject's characteristics. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 1. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

Scenario 4b

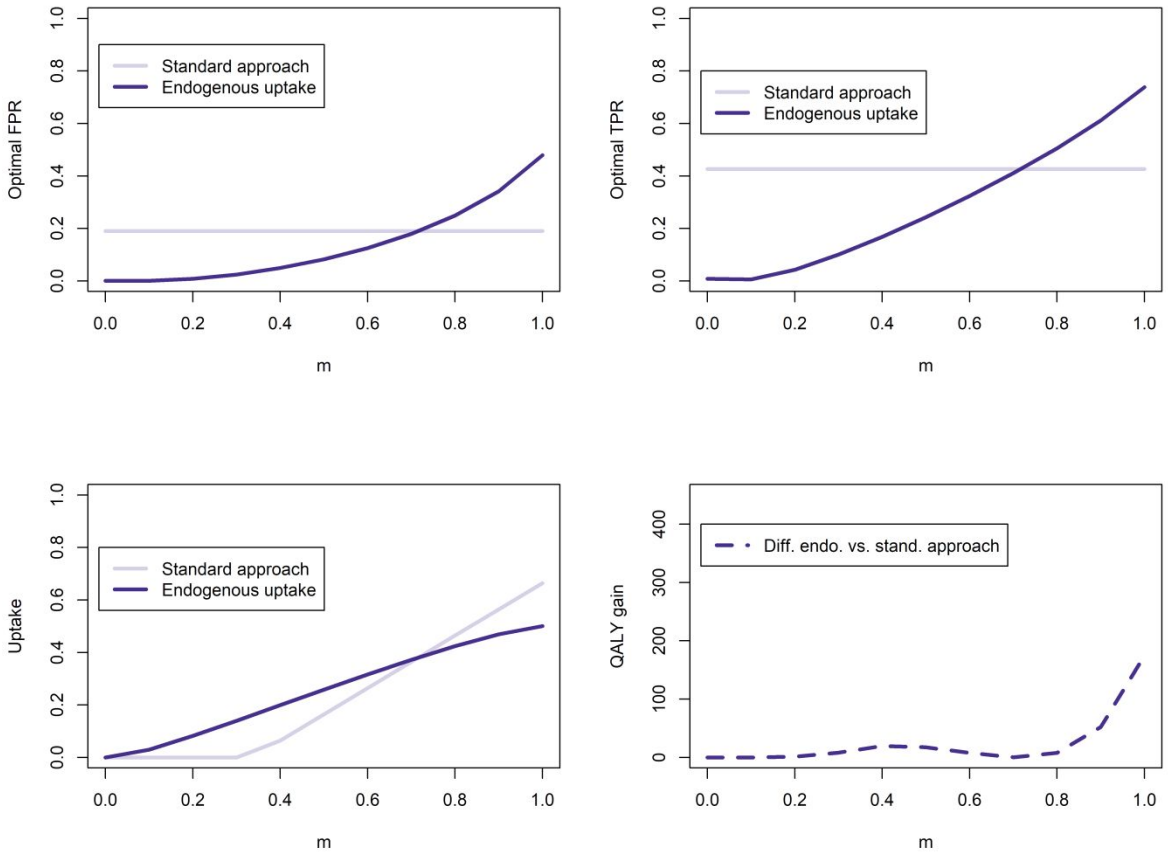


Figure A11. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The $(1-m)$% random uptake individuals are the same $(1-m)$% with the lowest $S^i$ in the target population. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 0. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

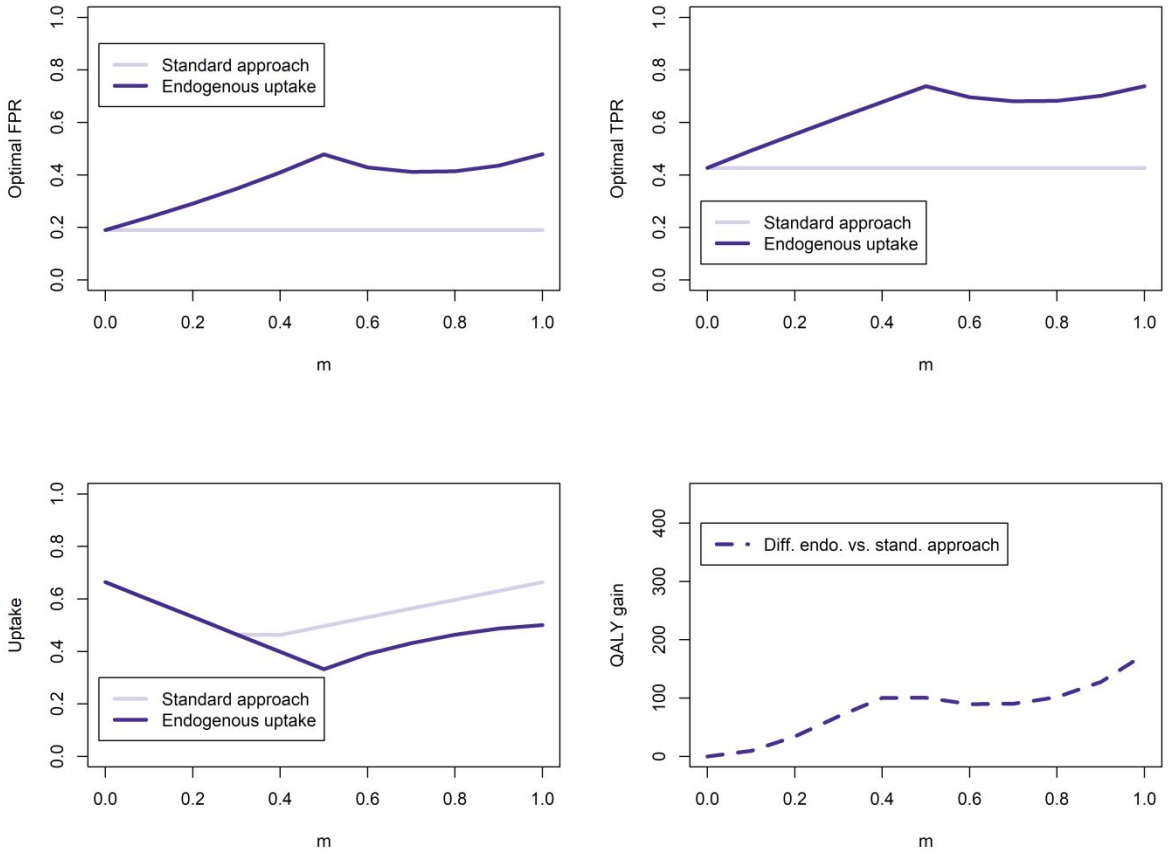


Figure A12. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The $(1-m)$% random uptake individuals are the same $(1-m)$% with the lowest $S^i$ in the target population. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 0.664. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

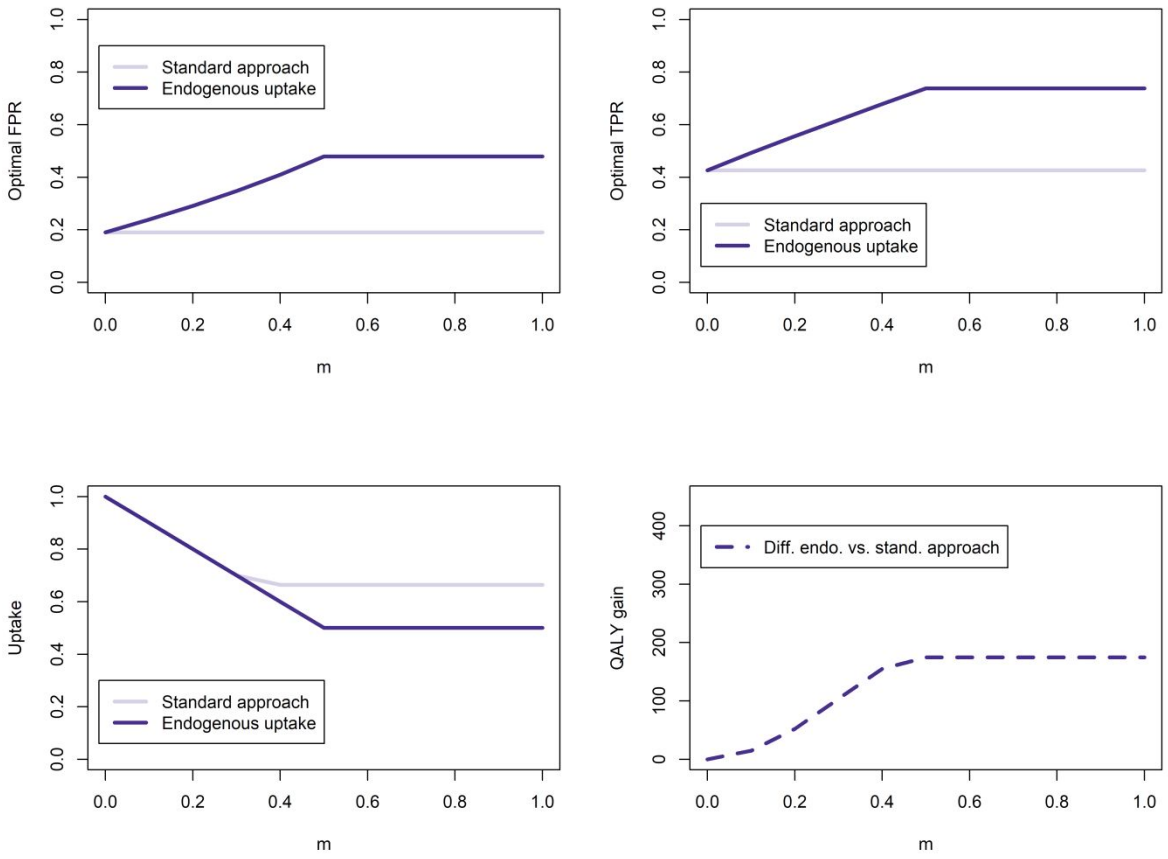


Figure A13. Comparison of the two approaches for different proportions of health utility maximisers ($m$). The $(1-m)$% random uptake individuals are the same $(1-m)$% with the lowest $S^i$ in the target population. Exogenous uptake rate of random uptake participants ($u^B$) is assumed 1. Assuming a target population of 1,000,000 subjects. ($m$ changed in steps of 0.1)

REFERENCES

1. Cooper JA, Parsons N, Stinton C, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. *British journal of cancer* 2018;118(2):285.
2. Dolan P. Modeling valuations for EuroQol health states. *Medical care* 1997:1095-108.
3. Moss S, Mathews C, Day T, et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. *Gut* 2017;66(9):1631-44.