Chapter 5

Design of stepwise screening for prediabetes and type 2 diabetes based on costs and cases detected

Published as:

de Graaf, G., Postmus, D., Bakker, S. J. L., & Buskens, E.
Design of stepwise screening for prediabetes and type 2 diabetes based on costs and cases detected. *Journal of Clinical Epidemiology, 68, 1010-1018* (2015)
ABSTRACT

OBJECTIVE To provide insight in the trade-off between cost per case detected and the detection rate in questionnaire based stepwise screening for impaired fasting glucose and undiagnosed type 2 diabetes.

STUDY DESIGN We considered a stepwise screening in which individuals whose risk score exceeds a predetermined cut-off value are invited for further blood glucose testing. Using individual patient data to determine questionnaire sensitivity and specificity and external sources to determine screening costs and patient response rates, we rolled-back a decision tree to estimate the cost per case detected and the detection rate for all possible cut-offs on the questionnaire.

RESULTS We found a U-shaped relation between cost per case detected and detection rate, with high costs per case detected at very low and very high detection rates. Changes in patient response rates had a large impact on both the detection rate and the cost per case detected, whereas screening costs and questionnaire accuracy mainly impacted the cost per case detected.

CONCLUSION Our applied method makes it possible to identify a range of efficient cut-offs where higher detection rates can be achieved at an additional cost per detected patient. This enables decision makers to choose an optimal cut-off based on their willingness to pay for additional detected patients.
INTRODUCTION

Type 2 diabetes mellitus (DM2) is a disease associated with a large burden at both patient and societal level. In Europe, an estimated 56.3 million people aged 20 – 79 have diabetes, of which 90% have DM2. The associated direct healthcare costs amounted to €111 billion in 2013. DM2 is therefore widely considered to be a major public health problem. There are two main strategies to address this issue. First is the reduction in the incidence of DM2 related complications through the early detection and treatment of asymptomatic DM2 patients (secondary prevention). Second is the provision of interventions aimed at slowing down the progression to DM2 in patients considered to be at high risk of developing DM2 (primary prevention), which is usually defined in terms of the presence or absence of prediabetes (i.e. impaired glucose tolerance or impaired fasting glucose (IFG)). As both strategies rely on blood glucose testing to either diagnose DM2 (secondary prevention) or to diagnose prediabetes and rule-out undiagnosed DM2 (primary prevention), a practical implementation of the second strategy results in finding undiagnosed DM2 patients as well. As a result, the combined screening for prediabetes and previously undiagnosed DM2 is more efficient and has gained widespread interest in the past years.

The target population for prediabetes and DM2 screening includes a large part of the entire population, but prevalences are low. Consequently, economic and logistic aspects of screening tools are an important consideration. To that end, blood glucose testing is generally considered too burdensome and costly to be applied in all individuals eligible for screening. Instead, consensus has been reached that screening should proceed in a stepwise manner by first making a preselection of high-risk individuals and then inviting those exceeding a predetermined threshold for further blood glucose testing. Risk questionnaires based on a small set of bio-characteristics have shown to be accurate predictors of DM2 risk, while being a relatively inexpensive form of
testing.\textsuperscript{10,11} Stepwise screening using a risk questionnaire has therefore found its way into several guidelines.\textsuperscript{12–14}

Although the strategy of stepwise screening is more feasible and practical, it inevitably also leads to a number of undiagnosed DM2 and prediabetes patients remaining undetected. Stepwise screening thus presents a tradeoff between feasibility, often measured in terms of the cost per case detected\textsuperscript{15–17}, and the detection rate (percentage of patients with disease in the target population that are detected through screening\textsuperscript{18}). In current guidelines, the selection of the cut-offs was based on an arbitrary value of absolute risk\textsuperscript{12,13} or was not supported at all\textsuperscript{14}. It therefore seems that the economic aspects were not explicitly considered during the formation of these guidelines, which may have been caused by the lack of insight in the trade-off between the cost per case detected and the detection rate.

In this paper, we seek to provide insight in the trade-off between cost per case detected and detection rate that comes with choosing a cut-off on a risk questionnaire. Furthermore, we want to estimate the effects of changes in patient response rates, screening costs, and questionnaire accuracy within the strategy of questionnaire based stepwise screening on this trade-off.

**METHODS**

**Structure of the stepwise screening program**

The stepwise screening program evaluated in this study was based on the Dutch guideline ‘Preventieconsult’.\textsuperscript{13} This guideline was developed to identify individuals at an increased risk for developing cardiovascular diseases, DM2, and kidney damage. We adapted this strategy to focus solely on IFG and DM2 by assuming the use of a dedicated DM2 questionnaire based on a version of the FINDRISC validated in the Dutch population.\textsuperscript{19}
The protocol for the screening program is as follows. Screening is initiated through the GP office by sending a questionnaire to all registered individuals of between the ages of 40 and 75 that have not been diagnosed with DM2 before. This questionnaire is returned to the GP office and assessed by the GP or a nurse. All individuals with a score equal to or above a predetermined cut-off value are invited for a consult and instructed to follow an 8-hour fasting protocol. At the consult, the answers of the questionnaire are verified and discussed and a fasting plasma glucose test is performed using a plasma calibrated capillary blood glucose meter. All patients with fasting plasma glucose levels of 6.1 mmol/L or higher are invited for a second consult and instructed to follow the fasting protocol again. During the second consult another fasting plasma glucose test is performed. The final diagnosis is based on the lower of the two test results. Thus, patients are diagnosed with DM2 if their fasting plasma glucose levels on both tests are 7.0 mmol/L or higher, with IFG if their fasting glucose level for the second test is between 6.1 mmol/L and 7.0 mmol/L, and with normal fasting glucose if their fasting plasma glucose level on the second test is below 6.1 mmol/L.

**Risk questionnaire**

The Dutch version of the FINDRISC questionnaire used in our screening design calculates a risk score based on five patient characteristics. These characteristics and the maximum number of points that can be acquired for each are: age (4 points), body mass index (3 points), waist circumference (4 points), the use of antihypertensives (2 points), and the occurrence of parental diabetes (5 points). This means that a patient can score between 0 points (lowest risk) and 18 points (highest risk). The original version of the Dutch FINDIRSC questionnaire includes an item on previously diagnosed DM2. As this was an exclusion criterion for the screening protocol as defined in the guideline, we removed this item from the questionnaire in our study.
Study population

The assessment of the stepwise screening program was performed using data from the PREVEND Groningen study, a cohort drawn from the general population in the city of Groningen in the Netherlands. Details of the study design have been published elsewhere. In short, a total of 40,856 individuals provided a urine sample and completed a questionnaire on demographics, history of cardiovascular and metabolic risk factors, known diabetes, medication use, and pregnancy. Pregnant females and patients using insulin were excluded. All participants with urinary albumin concentration of at least 10 mg/L willing to participate were enrolled in the study (n = 6,000). A random sample of those with urinary albumin concentration less than 10 mg/L were added to form a total study cohort of 8,592 participants. At baseline, participants underwent outpatient visits to assess demographics, anthropometric measurements, cardiovascular and metabolic risk factors, health behavior, and family history. Additionally, a blood sample was collected after an overnight fasting, from which fasting plasma glucose measurements were taken.

We selected all patients from the cohort that fulfilled the age criterion in the guideline (40–75 years, n = 6,244). For the purpose of this paper, participants in the cohort who were known to have diabetes (n = 149), who did not adhere to the fasting protocol (n = 857), or who did not have a baseline fasting plasma glucose measurement (n = 56) were excluded from the analysis. Known diabetes cases were identified either through the registered use of anti-diabetics in the pharmacy registry or the indication of being diabetic on the baseline questionnaire. Adherence to the fasting protocol was based on self-indication of consumption of any food or drinks other than water since midnight on the day of the glucose test. The use of antihypertensive medication was based on data in the pharmacy registry or self-reported use for those participants not present in the pharmacy registry. The PREVEND baseline questionnaire contained separate questions on the presence of DM2
Design of stepwise screening for type 2 diabetes

in both parents. These variables were combined into one variable indicating the presence of parental diabetes. Finally, all patients with missing data on one of the variables required to calculate the FINDRISC (i.e. age, sex, body mass index, waist circumference, use of hypertension medication, and parental diabetes) were excluded (n = 333).

**Outcome measures**

The performance of the questionnaire-based stepwise screening program was assessed using two outcome measures: 1) the number of IFG and DM2 patients identified as a result of screening as a percentage of the total number of IFG and DM2 patients in the target population (detection rate)\(^8\), and 2) the screening cost per case detected (CPCD). To compute these two outcome measures for a given cut-off on the FINDRISC, we first determined the expected cost and probability of being detected for each individual in our study population. These were calculated by rolling back the decision tree depicted in Figure 1. The individuals’ scores on the FINDRISC were computed based on the risk factor data taken from the baseline questionnaire of the PREVEND study. The detection rate was then obtained by summing all individual detection probabilities and dividing it by the total number of IFG and DM2 patients in the study population. The CPCD was calculated by summing all expected costs and dividing it by the sum of all individual detection probabilities. Confidence intervals (CI) for the outcome measures were obtained by repeating the analyses in 10,000 bootstrap samples from the study population and taking the 2.5\(^{th}\) and 97.5\(^{th}\) percentile of the outcomes.

**Base case analysis**

All cost and response parameters for the base case analysis are listed in Table 1. Probabilities of non-response at each step of the stepwise screening program were taken from the Preventieconsult trial\(^21\) and were assumed to be the same for all individuals. The Preventieconsult trial did not report screening
costs. Instead, we based these on a screening program for chlamydia that also send out invitations via mail.\textsuperscript{22} This program did however not involve the assessment of returned questionnaires. The total costs of questionnaire assessment were based on the costs for return postage (€0.50) and an estimation of the labor costs of evaluating the questionnaire (€0.50). Costs for a GP consult were taken from the Dutch reference price list.\textsuperscript{23} Lastly, a cost estimate for the glucose test was based on a commercial quotation.

**Table 1: Parameter data**

<table>
<thead>
<tr>
<th>Patient response rate</th>
<th>Base case scenario</th>
<th>Aphrodite scenario</th>
<th>Increased awareness scenario</th>
<th>Full response scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return of questionnaire (%)</td>
<td>75</td>
<td>55</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>Attend 1st consult (%)</td>
<td>72</td>
<td>73</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>Attend 2nd consult (%)</td>
<td>84</td>
<td>84</td>
<td>84</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Screening costs</th>
<th>Base case scenario</th>
<th>Double consult scenario</th>
<th>Nurse consult scenario</th>
<th>Email invitation scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitation and questionnaire (€)</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Questionnaire assessment (€)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Invitation for consult (€)</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Consult (€)</td>
<td>28.00</td>
<td>56.00</td>
<td>18.67</td>
<td>28.00</td>
</tr>
<tr>
<td>Fasting plasma glucose test (€)</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**Impact of selected parameters**

The impact of response rates, screening costs, and questionnaire accuracy on the outcome measures were calculated by changing those parameters in the decision tree model, rolling back the tree, and calculating the outcome measures using the same method described previously. Parameter values used in the impact scenarios for response rates and screening costs are shown in Table 1.
**Figure 1:** Model of the questionnaire based stepwise screening program. GP = General Practitioner; FPG = fasting plasma glucose; IFG = Impaired Fasting Glucose; DM2 = type 2 diabetes; TN = True Negative; FN = False Negative; TP = True Positive. Squares indicate decision notes, circles indicate chance nodes, and triangles indicate end notes of the decision tree. $p$ indicates the probability a patient enters that branch of the decision tree, and $c$ indicates the costs associated with entering that branch. $f(T,P)$ indicates that the probability for that branch is a function of the chosen threshold value as well patient characteristics, whereas $f(P)$ indicates that the probability is a function of patient characteristics only.


**Response rates**

The Dutch APHRODITE study, which also included questionnaires send out by GPs, found a response rate of 55% for the questionnaire and 73% for the first consult. A second response scenario was one where investment in public awareness would lead to an increase of 15% in response to both the questionnaire and the first consult, leading to response rates of 86% and 83%, respectively. To fully appreciate the effects of non-response on the outcomes of the screening, we also incorporated a hypothetical scenario where there is no non-response on all three invitations.

**Screening costs**

In the pilot study of the Preventieconsult, about half of the GPs indicated that they were unable to perform the consults within the standard duration of 10 minutes. This would mean that a 20-minute consult is required, which would double the consult costs. An alternative approach would be to have a nurse practitioner perform the consults. This would lead to an estimated reduction of the cost of the consults by one third. In a third and last scenario, the invitations and questionnaire would be sent out by email, reducing these costs to €0.00. The €0.50 for assessment of the questionnaire remains.

**Questionnaire accuracy**

A number of alternative questionnaires exist that could be applied in our study population, such as the Danish Diabetes Risk Score and the PM1 score. However, the difference in the accuracy of these screening tools with the FINDRISC is very small. An analysis of their impact would therefore not be very informative. Instead, we constructed three hypothetical questionnaires, with an area under the receiver operating characteristic (ROC) curve of approximately 5%, 10%, and 15% larger than the original FINDRISC questionnaire in our study population. For the questionnaires
with a 5% and 10% increased area under the curve, the FINDRISC score of all patients with IFG and DM2 was increased by 1 point, whereas the scores of all those with normal fasting glucose levels were lowered with one point. For the questionnaire with a 15% larger area under the curve the same approach was taken, but original scores were altered with 2 points. When this change led to questionnaire scores below 0 points or over the maximum of 18 points, these scores were set at 0 and 18, respectively. Subsequently, the ROC curves of these alternative questionnaires were inspected visually and additional changes were made to individual patient scores to improve the shape of the ROC and to approximate the intended area under the curve as closely as possible.

**Table 2**: Study population characteristics

<table>
<thead>
<tr>
<th></th>
<th>Normal fasting glucose</th>
<th>Impaired fasting glucose</th>
<th>Type 2 diabetes</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>4832</td>
<td>219</td>
<td>150</td>
<td>5201</td>
</tr>
<tr>
<td>Age, years (median (IQR))</td>
<td>52 (46 – 63)</td>
<td>60 (52 – 67)</td>
<td>63 (55 – 68)</td>
<td>53 (46 – 63)</td>
</tr>
<tr>
<td>Female sex</td>
<td>50.0%</td>
<td>38.4%</td>
<td>40.7%</td>
<td>49.2%</td>
</tr>
<tr>
<td>Fasting plasma glucose, mmol/L (median (IQR))</td>
<td>4.8 (4.4 – 5.1)</td>
<td>6.3 (6.2 – 6.6)</td>
<td>8.2 (7.3 – 10.9)</td>
<td>4.8 (4.5 – 5.3)</td>
</tr>
<tr>
<td>BMI kg/m2 (median (IQR))</td>
<td>25.9 (23.6 – 28.7)</td>
<td>28.7 (26.4 – 31.6)</td>
<td>28.9 (26.2 – 32.0)</td>
<td>26.1 (23.8 – 28.9)</td>
</tr>
<tr>
<td>Waist circumference, cm (mean ± SD)</td>
<td>89.9 ± 12.5</td>
<td>99.3 ± 11.4</td>
<td>100.4 ± 12.9</td>
<td>90.6 ± 12.7</td>
</tr>
<tr>
<td>Use of antihypertensive medication</td>
<td>16.2%</td>
<td>32.4%</td>
<td>38.7%</td>
<td>17.5%</td>
</tr>
<tr>
<td>Diabetes in family</td>
<td>15.8%</td>
<td>31.1%</td>
<td>25.3%</td>
<td>16.7%</td>
</tr>
<tr>
<td>FINDRISC questionnaire score (median (IQR))</td>
<td>7 (3 – 9)</td>
<td>10 (7 – 12)</td>
<td>10 (8 – 12)</td>
<td>7 (3 – 9)</td>
</tr>
</tbody>
</table>
RESULTS

Study population and questionnaire accuracy

The final sample of the study population consisted of 5,201 individuals, of which 219 (4.21%) had IFG and 150 (2.88%) had previously undetected DM2, yielding in a total number of 369 cases (Table 2). Compared to those with normal fasting glucose, patients with IFG and DM2 are on average older, more often male, have a higher BMI, larger waist circumference, use antihypertensive medication more often, and more frequently have family members with diabetes (Table 2). Within this study population, the FINDRISC based questionnaire had an area under the ROC curve of 74.3% (95% CI 71.9%-76.7%) (Figure 2 left panel). The area under the ROC curve for the IFG patients alone was 72.5% (95% CI 69.2%-75.7%), whereas that for the DM2 patients alone was 74.8% (95% CI 71.3%-78.2%) (results not shown). Sensitivity and specificity for each cut of point of the FINDRISC for the IFG and DM2 patients separately are shown in Table 3. The three hypothetical questionnaires used in the sensitivity analysis of questionnaire accuracy had areas under the curve of 80%, 85% and 90% (Figure 2 right panel).

Base case analysis

A U-shaped relation was found between detection rate and CPCD, with a high CPCD at both very low and high detection rates (Figure 3 top left panel). Lowest CPCD was achieved at cut-off 10, being €445 (95% CI €398-€507) with a detection rate of 24.6% (95% CI 21.3%-28.0%). Increasing the cut-off to 11 led to unfavorable effects on both outcomes, as CPCD increased to €481 (95% CI €423-€555) and the detection rate decreased to 19.3% (95% CI 16.3%-22.4%). On the other hand, decreasing the cut-off below 10 had a favorable effect on the detection rate, but an unfavorable effect on the CPCD. These cut-offs therefore present a trade-off between higher detection rates and higher CPCD. Cut-offs below 6 resulted in very little additional detection, but did lead to marked increases in CPCD. Due to the non-response, the detection
The subgroup analysis demonstrated that the detection rate at each cut-off point is very similar in both IFG and DM2 patients (Table 3). This means that the proportion of IFG and DM2 patients identified at each cut-off is similar to proportion of prevalences in the target population.

**Impact of selected parameters**

The two outcome measures were impacted differently by the three parameters in the impact analysis (Figure 3). Changes in response levels had a large impact on both outcome measures. An increase in response led to a higher detection rate and lower CPCD and vice versa (Figure 3 top right panel). In the scenario with full response on all three invitations, the detection rate reached 100% (cut-off 0) and the detection rate at each cut-off corresponded to the sensitivity of that cut-off as shown on the ROC curve (Figure 2 left panel). In contrast, changes in the screening cost only influenced the CPCD. Changes

---

**Figure 2:** Receiver operator characteristic curves of the base case (left) and impact analysis (right). Numbers accompanying points on the curve indicate the cut-off score on the questionnaire. Numbers between parentheses are the 95% confidence interval of the area under the curve.
Chapter 5

Figure 3: Outcomes of the base case (top left), impact analysis of patient response rates (top right), impact analysis of the screening costs (bottom left), and impact of the questionnaire accuracy (bottom right). Numbers indicate the cut-off score on the questionnaire. Parameter values of the impact analysis scenarios (gray curves) are shown in Table 1.
in costs later in the screening process (consult costs) had a larger impact on
the lower cut-offs, whereas changes in costs early in the screening process
(invitations and questionnaire assessment) had a larger impact on the higher
cut-offs (Figure 3 bottom left panel). A change in accuracy did change both
the CPCD and detection rate at each individual cut-off, but did not change
the maximum achievable uptake. Finally, a larger area under the curve led to
both a higher detection rate and a lower CPCD (Figure 3 bottom right panel).
The effects of increased accuracy were most prominent in the cut-offs that
are closest to the upper left corner in the ROC, whereas the position of the
extremes remained more or less the same.

**Table 3**: Subgroup analysis for questionnaire accuracy and detection rate

<table>
<thead>
<tr>
<th>Cut-off</th>
<th>Impaired fasting glucose patients</th>
<th>Type 2 diabetes patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>99.1</td>
<td>7.2</td>
</tr>
<tr>
<td>2</td>
<td>98.6</td>
<td>9.4</td>
</tr>
<tr>
<td>3</td>
<td>96.8</td>
<td>19.3</td>
</tr>
<tr>
<td>4</td>
<td>94.5</td>
<td>27.0</td>
</tr>
<tr>
<td>5</td>
<td>90.4</td>
<td>34.5</td>
</tr>
<tr>
<td>6</td>
<td>87.7</td>
<td>40.2</td>
</tr>
<tr>
<td>7</td>
<td>80.4</td>
<td>48.3</td>
</tr>
<tr>
<td>8</td>
<td>72.1</td>
<td>59.3</td>
</tr>
<tr>
<td>9</td>
<td>66.7</td>
<td>67.8</td>
</tr>
<tr>
<td>10</td>
<td>54.3</td>
<td>78.2</td>
</tr>
<tr>
<td>11</td>
<td>43.4</td>
<td>84.2</td>
</tr>
<tr>
<td>12</td>
<td>29.2</td>
<td>90.4</td>
</tr>
<tr>
<td>13</td>
<td>20.5</td>
<td>93.6</td>
</tr>
<tr>
<td>14</td>
<td>14.2</td>
<td>96.1</td>
</tr>
<tr>
<td>15</td>
<td>8.2</td>
<td>97.7</td>
</tr>
<tr>
<td>16</td>
<td>5.0</td>
<td>98.7</td>
</tr>
<tr>
<td>17</td>
<td>3.7</td>
<td>99.5</td>
</tr>
<tr>
<td>18</td>
<td>0.9</td>
<td>99.8</td>
</tr>
</tbody>
</table>
All three parameters had an impact on the trade-off between detection rate and CPCD. This trade-off became more favorable (i.e., a higher detection rate can be achieved for a smaller increase in CPCD) when response rates increased, late-stage screening costs decreased (nurse consult), or accuracy of the questionnaire was reduced. However, the latter also resulted in lower absolute detection rates for any given cut-off.

The cut-off with the lowest CPCD changed in some of the impact scenarios. The largest shift occurred in the email invitation scenario, where it shifted to 12. Contrarily, in the scenario with 33% lower consult costs the cut-off with lowest CPCD decreased to 9. This was the same in the scenario with reduced patient response rates. Lastly, in the scenario with the largest increase in accuracy (area under the curve 90%), the cut-off with lowest CPCD increased to 11.

**DISCUSSION**

In this study, we set out to provide insight in the trade-off between CPCD and detection rate when choosing a cut-off on a risk questionnaire used in stepwise screening for IFG and DM2. At low cut-off scores, unnecessary GP consults and glucose tests are provided to a large number of false positive patients. Contrarily, at high cut-off scores the initial costs of sending out invitations and questionnaires are shared by a small group of detected cases due to the large number of false negative patients. Combined, these aspects resulted in a U-shaped relation between CPCD and detection rate with the lowest CPCD attained at cut-off 10. Additionally, we investigated the impact of possible changes within the framework of questionnaire based stepwise screening on the trade-off between CPCD and detection rate. Changes in patient response rates had the largest impact on the results of screening as these had a strong effect on both detection rate and CPCD. Changes in screening costs or accuracy mainly affected CPCD. In terms of CPCD, the effect of 15% more response was very similar to a reduction of consult costs by 33% for the cut-off.
with lowest CPCD. However, in the increased response scenario the detection rate additionally increased by one third at this cut-off.

Increasing the cut-off from the score with the lowest CPCD would result in a lower detection rate and higher CPCD. These cut-offs can therefore be discarded as sub-optimal. All remaining cut-offs, from the lowest up to and including the one with the lowest CPCD, present a trade-off where additional detection can be gained for an increase in CPCD. However, because positive patients with a very low score are very rare, decreasing the cut-off in the lower range of cut-offs results in very little additional detection but a large increase in CPCD. It is therefore possible to identify a range of efficient cut-offs, in our case from 6 up to and including 10. Within this set, decision makers would have to determine their willingness to pay for the additional detection in order to find the optimal cut-off.

One approach to find the optimal cut-off would be to define a maximum CPCD based on an investment perspective on screening. Taking such a perspective, the maximum CPCD is determined by the average gain in net monetary benefit that can be achieved by treating the detected cases. If all cases were to have the same level of utility in being detected, the maximum CPCD would form a horizontal line in the plot of CPCD versus detection rate. In reality, however, the impact of treatment on the downstream costs and health effects differs between patients depending on their age, disease status (IFG or DM2), whether they have comorbidities, and on other patient characteristics, meaning that the maximum CPCD is likely to vary with the cut-off selected. If the entire U-curve is above the maximum CPCD curve, stepwise screening is not viable from a health economic perspective. In contrast, if the maximum CPCD curve crosses the U-curve or if the entire U-curve is below the maximum CPCD curve, the rightmost cut-off under the maximum CPCD curve leads to the largest feasible uptake. In these situations, there are two ways to approach an optimal cut-off. A cost-dominated approach would start at the cut-off with the lowest CPCD and require the decision
maker to determine whether he is willing to pay the additional CPCD for the additional cases detected when decreasing the cut-off by one point. This is then repeated until the willingness to pay is not large enough to decrease the cut-off more. Alternatively, a detection-dominated approach would start at the rightmost cut-off below the maximum CPCD curve. The cut-off is then increased as long as the reduction in CPCD is large enough to offset the reduced detection (willingness to forgo). A different perspective a decision maker could have is one of a budget constraint, in which there is a maximum amount of resources available to be allocated to screening. Taking such a perspective, the optimal cut-off can be found by calculating the total costs for screening, which, although not displayed in our analyses, can easily be obtained from the presented model. The decision maker can then simply look for the lowest cut-off for which the total costs of screening is still within his budget.

Economic considerations are important when initiating a large screening program as there will always be a considerable budget impact. Despite of this, CPCD is not yet routinely considered when evaluating different screening designs. One recent study compared a large number of different single step and stepwise screening strategies on their detection rate, false positive rate, and CPCD. The authors concluded that stepwise screening using available data in GP records or self-administered questionnaires were the most feasible strategies in terms of CPCD. However, they only considered a limited number of cut-offs as alternative scenarios, and attendance rates were fully based on assumptions. The only other studies known to us to use CPCD as an outcome measure did not consider stepwise screening using a questionnaire, but rather two steps of blood glucose testing. In one of these, the authors applied a similar approach to ours by assessing the detection rate and CPCD for all possible cut-off points. However, rather than summarizing the performance on these two outcomes in a single graphical presentation as we did in this study, they presented the results separately for the detection rate and CPCD, and declared the cut-off with the lowest CPCD as the optimal one. As
detecting patients in the target population is the primary aim of screening, it is highly unlikely that policy makers would indeed consider the point with lowest CPCD as optimal.

Although we presented results of a specific Dutch scenario in terms of costs and prevalences, the main findings hold for any type of stepwise screening in any target population. The results of our study therefore have a broad set of implications going beyond the realm of stepwise screening for prediabetes and DM2. First and foremost is the conclusion that selecting a cut-off without evaluating its implications in terms of CPCD and detection rate could lead to a waste of resources and missed opportunities of patients who could be treated (when a cut-off outside the efficient range is selected). Alternatively, it could lead to a possibly undesirable implicit statement about the willingness to pay for additional cases detected. Second, innovations in information technology infrastructure make it possible to do more with patient data in GP records. There is a move towards automated risk profiling within GP administration systems. This changes the costs associated with the first steps of screening, similar to that in our ‘email invite’ cost scenario. We have demonstrated that this change has a large impact on the range of efficient cut-offs, and it is therefore advisable to revisit the cut-off decision when such changes take effect. Finally, our impact analysis provides insight in where further improvements within the strategy of stepwise screening would have the most beneficial effects on the trade-off between CPCD and detection rates. In the past, research has mainly focused on improving the accuracy of risk questionnaires through the identification of novel biomarkers that would further improve risk stratification, while there has been little focus on implementation aspects such as response rates. However, our analysis indicates that improving the latter has a much stronger effect on the outcomes of stepwise screening than improving the former. This suggests that funds available for research would have a higher impact when they are invested in implementation research, rather than in laboratory research.
In conclusion, policy makers should be aware that the choice of a cut-off on the first test in a stepwise screening program has direct implications on the feasibility, effectiveness, and budget impact of the program. This choice should therefore be based on an integral assessment of all these aspects rather than solely on test accuracy. The methods presented in this paper can assist policy makers to do so.
REFERENCES


Design of stepwise screening for type 2 diabetes