Quality of life assumptions determine which cervical cancer screening strategies are cost-effective

Inge M.C.M. de Kok, PhD [a]*, i.dekok@erasmusmc.nl
Ida J. Korfage, PhD [a]*, i.korfage@erasmusmc.nl
Wilbert B. van den Hout, PhD [b], w.b.van_den_hout@lumc.nl
Theo J.M. Helmerhorst, MD PhD [c], t.helmerhorst@erasmusmc.nl
J. Dik F. Habbema, PhD [a], j.d.f.habbema@erasmusmc.nl
Marie-Louise Essink-Bot, MD PhD [d†]
Marjolein van Ballegooijen, MD PhD [a], m.vanballegooijen@erasmusmc.nl

[a] Erasmus University Medical Center, Department of Public Health, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands
[b] Leiden University Medical Center, Department of Medical Decision Making, P.O. Box 9600, 2300 RC Leiden, the Netherlands
[c] Erasmus University Medical Center, Department of Gynaecology and Obstetrics, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands
[d] University of Amsterdam, Academic Medical Center, Department of Public Health, PO Box 22660, 1100 DD Amsterdam, the Netherlands.
* This concerns a shared first authorship, because both authors contributed equally

Novelty and Impact

This is the first study that empirically obtained utility scores for quality of life associated with cervical cancer screening and treatment in women in all relevant health states, ranging from being invited for screening to being disease-free after primary treatment of cervical cancer or having advanced cancer. Furthermore, we show that quality of life assumptions determine the cost-effectiveness of cervical cancer screening and that the measure used to empirically assess utilities is crucial for cost-effectiveness conclusions.

Funding: The Dutch Cancer Society (EMCR 2003–2775)

Corresponding author:
Inge M.C.M. de Kok
Erasmus MC, Dept. of Public Health
P.O. Box 2040, 3000 CA Rotterdam, the Netherlands
Tel. no.: +31 10 704 3721
Fax no.: +31 10 703 8475
E-mail: i.dekok@erasmusmc.nl

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ijc.31265

ABSTRACT

Quality adjusted life years are used in cost-effectiveness analyses (CEAs). To calculate QALYs, a 'utility' (0-1) is used for each health state induced or prevented by the intervention. We aimed to estimate the impact of quality-of-life (QoL) assumptions (utilities and durations of health states) on CEAs of cervical cancer screening. To do so, twelve alternative sets of utility assumptions were retrieved from published cervical cancer screening CEAs. Two additional sets were based on empirical QoL data that were integrally obtained through two different measures (SF-6D and EQ-5D) from eight groups of women (total n= 3,087), from invitation for screening to diagnosis with cervical cancer. Per utility set we calculated the number of quality-adjusted days lost (QADL) for each relevant health state in cervical cancer screening, by multiplying the study-specific assumed disutilities (i.e. 1-utility) with study-specific durations of the loss in QoL, resulting in 14 'QADL-sets'. With microsimulation model MISCAN we calculated cost-effectiveness of 342 alternative screening programs (varying in primary screening test [Human Papillomavirus (HPV) versus cytology], starting ages, and screening interval) for each of the 14 QADL-sets. Utilities used in CEAs appeared to differ largely. We found that ten QADL-sets from the literature resulted in HPV and two in cytology as preferred primary test. The SF-6D empirical QADL-set resulted in cytology and the EQ-5D one in HPV as preferred primary test. In conclusion, assumed utilities and health state durations determine cost-effectiveness of cervical cancer screening. Also, the measure used to empirically assess utilities can be crucial for CEA conclusions.

INTRODUCTION

Cervical cancer screening, still widely based on cytological Pap test, decreases mortality and incidence from this cancer.[1] However, changes are ongoing in cervical cancer prevention, such as the introduction of human papilloma virus (HPV) screening and HPV vaccination programs. An important criterion for appraising the validity of a new screening programme is that its benefit (i.e. preventing clinical cancer and cancer death) should outweigh its physical, psychological, and societal harm (caused by the test and screening-related diagnostic procedures and treatment).[2] This is assessed in a cost-effectiveness analysis (CEA). CEAs can use 'quality adjusted life years' (QALYs) as the summary effect measure. The QALY is a generic effect variable of disease burden, including both the quality and the length of life lived.[3] To calculate QALYs, a quality-adjustment weight or 'utility' is applied to each relevant health state. A utility is a number anchored at 0 (dead) and 1 (perfect health), indicating the preferability of that health state in terms of quality of life. Because of their crucial impact on CEAs, and thus on health care decision making about the introduction of e.g. screening programs or the choice about specific screening tests, reliable estimates of the utility scores of all health states induced and prevented by screening are important. For cervical cancer screening, after early detection via cytological abnormalities, women with cervical intraepithelial neoplasia (CIN) can be treated so that potential development of these abnormalities into cancer is prevented. The downsides of cervical cancer screening are lack of specificity, overdiagnosis and overtreatment, since most of the abnormalities found in the screening and follow up will never develop into clinical cervical cancer. Notably, it was shown that in the Netherlands one prevented cervical cancer death due to screening entails an estimated 2,097 women being screened, 64 women being sent to triage, and 42 women being referred for colposcopy, with punch biopsy and eventually treatment of mostly preinvasive neoplastic conditions.[4] The numbers of women needed to be screened or sent to triage or colposcopy to prevent one cervical cancer death increase by implementation of a more sensitive (but less specific) screening test. Recently it was shown, that the implementation of primary HPV screening in the Netherlands may lead to a threefold increase in the number of (false-positive) referrals to the gynaecologist.[5] Women

screened and referred to a gynaecologist may experience distress and anxiety.[6-8] Per woman this may have a limited effect, but considering the large numbers of women screened and referred, small quality of life losses for these harms might still result in relevant effects at population level. We currently lack empirically derived utility estimates for all relevant health states in cervical cancer screening. Currently, available utility estimates in the literature are conflicting, which results in parameter uncertainty.[9]

The present study first aims to estimate the impact of the variation in currently used sets of utility assumptions on the estimated cost-effectiveness of cervical cancer screening programs. To do so, we will review utility assumptions in the CEA literature. Next, we will use microsimulation modelling to estimate the impact of the differences in utility assumptions on the cost-effectiveness of cervical cancer screening (in a population unvaccinated against HPV) and on the preferred screening strategies.
Second, we derived two sets of utilities based on a state of the art empirical study. In this study, we have performed an integrated questionnaire study addressing all states related to the cervical cancer-screening programme in a generic and standardized way, resulting in two types of utilities for all relevant health states. Selected parts of this empirical study have been published earlier [6, 10-12] The study results will be presented here as a whole. Next, we will evaluate how use of our empirical utility measurements affects the microsimulation cost-effectiveness results, and how this compares to using utility estimates applied in the cervical cancer screening CEA literature.

METHODS

Part 1: Literature review

To collect the utilities that have been used so far, we reviewed the literature for CEAs of cervical cancer screening published between 2003 and 2015. We used the following search terms in Pubmed: cost-benefit analysis; cost-effectiveness analysis; uterine cervical neoplasms; cervical cancer; screening; early detection of cancer; quality-adjusted life years; QALY. We selected papers that examined cost-effectiveness of screening from Europe, North-America and Australia that are published in English. We identified 19 studies that used QALYs.[4, 13-30] Two pairs of studies used the same set of utility assumptions (i.e. same model and research group), and each of these pairs were combined ([4]and [29]; [17] and [18]). One study was excluded since it also used a common set of utility assumptions that was already included.[23] One study was excluded since the utilities used were not reported.[25] The resulting 15 studies defined different health states in cervical cancer screening. We first summarized the published health states into 13 health states (primary screening; positive primary screentest; false positive test; referral for triage test; referral for colposcopy; false positive referral; CIN1; CIN2; CIN3; FIGO 1; FIGO2+; Survivors; Palliative phase). We then for each study assessed the number of days lost due to diminished QoL ('quality-adjusted days lost' (QADL)) for these different health states, by multiplying the study-specific assumed disutilities (i.e. 1-utility) with the study-specific mean durations of the loss (or gain) in QoL. Three studies[20, 28, 30] were excluded because of absence of data on the durations. For five studies[4, 16, 24, 26, 27] we had to make assumptions (based on limited available information) on durations (see footnotes Table 2). So, the literature review resulted in 12 different 'QADL-sets'. Appendix Table 3 shows the assumed utilities in the twelve included published CEAs per health stage, as well as the source of each of the assumed utilities.

Part 2: Questionnaire study

We empirically obtained utility scores, indicating quality of life, for 7 study groups of women: invited for cervical cancer screening (n=1,023 respondents, of whom 905 participated in screening, response 60%); having borderline or mildly dyskaryotic (BMD) pap

test results (n=270; response 49%); referred for colposcopy (n=132); treated for a precursor of cervical cancer (n=81; response 49%); diagnosed with cervical cancer (n=77); disease-free after primary treatment of cervical cancer (n=285; response 69%), and a reference group (n=835, response 46%, Appendix Table 2). The methods and results have been published in detail for the EQ-5D scores of the screening participants,[10] the group with borderline or mildly dyskaryotic pap test results,[6] women referred for colposcopy,[12] and women who are disease-free after treatment.[11] Data were collected by self-administered questionnaires, containing both the EQ-5D and SF-6D questions. At the time of data-collection HPV screening had not been introduced yet. A description of the data-collection procedures and the background characteristics of the participants is included in the Appendix (Appendix Tables 1 and 2). No women with advanced cervical cancer (i.e. palliative phase) could be included due to logistic reasons; for this group we used utility estimates of women with breast or lung cancer.[31, 32] The ethics review committee of Erasmus University Medical Center Rotterdam approved the research protocol.

Utility measures

A quality-adjustment weight or 'utility' is a number anchored at 0 and 1, with "perfect health" carrying a weight of 1 and dead carrying a weight of 0. We used both the SF-6D and the EQ-5D. The SF-6D, based on a subset of SF-12 responses,[33] is composed of six dimensions: physical functioning, role limitations, social functioning, bodily pain, mental health, and vitality. It was derived from a valuation study among a representative sample of the general public in the UK, using the standard gamble valuation technique.[33] The EQ-5D classification consists of five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[34] Response categories indicate 1) no problems; 2) some problems; or 3) serious problems. Responses were linked to utility scores as obtained in the UK using time trade-off methodology.[34]Information on age, marital status, education, and profession of respondents was obtained through the questionnaires.

Statistical analyses

Paired t-tests were used to assess the statistical significance of differences between SF-6D and EQ-5D scores. Pearson product-moment correlations were used to calculate within-person correlations between SF-6D and EQ-5D assessments. All analyses were performed in SPSS, version 20.

Calculation of quality adjusted days lost (QADL)

Based on the questionnaire study we calculated the number of QADL for those health states for which we empirically found a significant different utility (measured by the EQ-5D and the SF-6D, separately) compared to the reference population (screening ages in case of screening health states and all ages in case of cancer health states). For these health states we calculated the disutility by subtracting the measured utility (the point estimate) in that specific health state from the utility measured in the general population (screening ages in case of screening stages). In case of 'terminal care' we subtracted the utility in breast cancer patients in the last year of life from the utility in the general population, both measured by the EQ-5D. We then calculated the number of QADL for the different health states, by multiplying the measured disutilities with the durations of the loss in QoL per state. The durations where based on 1) the Dutch screening guideline in case of a BMD result (i.e. 15 months), 2) measured duration of significant QoL loss in case of longitudinal measurements for referral to the gynaecologist and cancer health states, 3) 10 years in case of cancer survivor based on the definition used in the questionnaire study, 4) last year of life in case of terminal care (i.e. 1 year).

Part 3: Microsimulation modelling

We used the microsimulation screening analysis (MISCAN) model[29] to estimate the costs and QALYs gained for alternative cervical cancer screening programs, in a Dutch (unvaccinated against HPV) population. The model and the inputs used are described in the Appendix, as well as in an earlier publication.[29] Also, the calibration and validation of the model are described in an online technical appendix [http://hdl.handle.net/1765/31582].

We simulated 1,000,000 unvaccinated women born between 1939 and 1992. The alternative strategies included primary cytology and primary HPV (followed by cytology triage) screening, with respective triage strategies that according to the MISCAN model used were cost-effective.[29] We considered all screening policies with starting ages of 25, 27, 30 or 32 years that comprise at least three and at most ten screenings in a woman's lifetime. Policies had an interval of at least three years and at most ten years, policies that include screenings over the age of 70 years were excluded. This resulted in 342 simulated screening policies. Based on monitoring (program), trial and administrative (program and hospital) data we made assumptions on screening attendance, test characteristics and costs (see Appendix Table 4). The sensitivity of the HPV test (the probability of a positive test result if an HPV infection is present) was estimated at 94%, and the sensitivity of cytology was assumed to be 40% for CIN 1, 50% for CIN 2, and 75% for CIN 3 and invasive cervical cancer (see Appendix Table 4). The specificity of the HPV test (probability of a negative test for women without high-risk HPV infections) is assumed to be 100%, and the specificity of cytology (probability of a negative test for women without CIN or cancer) is estimated to be 98.5%. Lack of specificity in case of HPV screening was accounted for by the inclusion of fast-clearing HPV infections.

The lifelong costs and effects of each simulated screening program were counted for the period from 2011 onwards (until all women have died), and discounted at an annual rate of 3% towards the year 2011. For each woman, we calculated the number of QALYs as the weighted sum of the number of years spent in each of the health states, using the state specific utility weights. The total effectiveness of screening (QALYs gained) is determined as the difference in the number of QALYs between the situation with screening and the situation without screening. We used a similar approach to determine the net costs of screening.

Programs that were more costly and less effective than other programs were ruled out as non-efficient (by both simple and extended dominance). The remaining programs constitute the frontier of efficient screening programs, see the lines in Figure 2. For each 'QADL-set' we determined which screening programs were located on the cost-effectiveness frontier.

Based on the incremental cost-effectiveness ratios (ICERs) between these programs we determined which screening program was preferred considering cost-effectiveness thresholds of €20,000 and €50,000 per QALY gained.

RESULTS

Part 1: Literature review

Both for screening and treatment phases we found a large variation in number of QADL between the published studies (Table 1). For example, the number of QADL for a cervical intraepithelial neoplasia (CIN) 2 or CIN 3 lesions varied from 0 to 47.5. For a FIGO (International Federation of Gynecology and Obstetrics) stage 2+ cancer, the number of QADL varied from 51 to 1460, which is an 80-fold difference. This variation is the result of differences in utility values (see Appendix Table 3), as well as differences in durations. We found that all CEAs ultimately refer to 12 different original utility studies, published between 1991 and 2012 (see Appendix Table 4).

Part 2: Questionnaire study

SF-6D utility scores

The mean utility score was 0.83 (confidence interval (CI): 0.82-0.84) in the general adult female population and 0.84 (CI: 0.83-0.85) in the on average somewhat younger women in screening ages (30-60 years), see Table 2 and Figure 1A. Screen participants' mean scores were slightly higher: 0.85 before and after screening and 0.86 after receiving the test results. The mean utility of women who were recommended to have a repeat Pap test because of borderline or mild dyskaryotic test results were lower than those of the reference population and the screening invitees. Before onset of treatment, both FIGO groups reported similar utility scores. The FIGO 1 group (n=47) improved steadily from 0.73 at baseline to 0.81 one year later, while the FIGO 2+ group (n=16) reported 0.71 at baseline, 0.63 at three months follow-up, and 0.71 and 0.75 at six and twelve months follow-up. The mean utility of tumorfree cancer survivors was 0.80, and thus comparable to that of women with cervical cancer FIGO 1 at 1-year post –diagnosis.

Overall, SF-6D utility scores were best (0.86, CI: 0.85-0.87) in screen participants after receipt of Pap test results and worst in women with cervical cancer (FIGO 2+) at three months after diagnosis (0.63, CI:0.57-0.70).

Comparison of SF-6D and EQ-5D utility scores

SF-6D and EQ-5D utility scores were always significantly correlated and the patterns of SF-6D and EQ-5D utility scores were roughly similar (Figures 1.A and 1.B). On average, SF-6D utility scores were lower than EQ-5D utility scores, but in more serious health states, such as diagnosis with cervical cancer, SF-6D and EQ-5D scores no longer significantly differed.

Number of QADL

The number of QADL based on our questionnaire study differ substantially from those used in the CEAs obtained from the literature. The QADL, calculated with the SF-6D, due to having gone through a triage episode is 14 days. With the EQ-5D, no significant different utility was measured for women in triage compared to the references population (screening ages) (Table 2), so no QADL were measured in the screening stages. The QADL (SF-6D) due to a FIGO2+ cancer diagnosis is 161 (51+110) days for survivors and 234 (51+183) days for those who die from it (Table 2). In case of the EQ-5D, these figures are 166 (21+146) and 204 (21+183), respectively.

Part 3: Microsimulation modelling

Per 'QADL-set' different screening programs are presented on the cost-efficient frontier (Figure 2). Table 3 shows per QADL-set the preferred screening strategy at cost-effectiveness thresholds of €20,000 and €50,000 per QALY gained, according to the MISCAN model. At a willingness to pay of €20,000 per QALY, eleven QADL-sets (ten from the literature and the EQ-5D empirical data) preferred HPV screening, all starting at age 30 or 32, performing 3 or 4 tests with a 5 or 6 year interval. Our empirical SF-6D data and two other sets from the literature preferred cytology screening, starting at age 30 or 32, performing 4 or 5 tests at a 4, 5 or 6 year interval.

At €50,000 per QALY, the optimal strategies tended to have lower starting ages (age 27 to 30), with more tests (7 to 10) and shorter intervals (4 to 6 years). However, the preferred screening tests for each QADL-set did not change. The QADL-set 'Kitchener (Simonella)'[22]

resulted in less frequent screening, since this set showed a considerable loss in QoL related to attending screening.

DISCUSSION

We showed that differences in quality of life assumptions as they are found in published cervical cancer screening CEAs and as measured in an integral questionnaire study using utility measures, lead to different conclusions about cost-effectiveness of strategies. Differences in quality of life assumptions result in different number of QALYs gained for the same screening strategy and, as a result, different ICERs between screening strategies. In general, it largely depends on utility losses assumed for positive screening results including the associated follow up, which screening modality is preferred. Primary HPV screening, with relatively more positive screen test results than cytology screening, is preferred in case of lower utility loss assumptions for the screening phase (used in ten studies), and cytology in case of higher utility losses (used in two studies). For example, the assumed number of QADL due to referral in the study of Karnon et al. [24] is significantly higher than in the other scenarios, and therefore having a (false-positive/clinically not relevant) referral for colposcopy has more impact in this scenario than in the other scenarios. In this example, one could argue that the fact that a large majority of studies used lower utility losses in the screening phases, provides some strong certainty that HPV screening is to be preferred. In a situation with high uncertainty, however, such counting of studies not necessarily leads to the correct conclusions. Clearly, focussing further screening related QoL research on follow up after positive screen results is important.

The use of the presented empirical utility data measured by the SF-6D resulted in the preference for primary cytology screening, whereas in most utility sets used in the CEA literature, primary HPV screening was preferred. This was caused by the fact that relatively larger quality of life losses were found with the SF-6D for triage and referral for colposcopy, which is a health state that will occur more frequently with primary HPV screening than with cytology.[35] As a result, more quality of life will be lost due to the screening itself with primary HPV screening compared to cytology and therefore primary cytology becomes the preferred strategy. Our finding that quality of life assumptions influence the results of CEA of cervical cancer prevention is in line with previous studies.[22, 36, 37]

We found large differences in utility estimates and durations used in CEAs. It appeared that none of these CEAs was based on utility estimates derived from empirical data of cervical cancer screenees or actual patients who were faced with the studied diagnosis in real life, but on data collected in other (patient) groups (i.e. patients with other conditions than (precursors of) cervical cancer). However, patients (that experience the studied disease) often report better quality of life than healthy people who are asked to imagine the patient's circumstances; the so-called disability paradox.[38] Causes of this paradox include loss aversion, focusing illusion, and underestimation of their own adaptation in healthy people and adaptation processes in patients.[39-41] These phenomena indicate the importance of empirically derived utility estimates.

In this first study to empirically assess the QoL effects of all relevant stages in population-based organized cervical cancer screening, we observed, per woman involved, only limited QoL loss due to screening as long as there was no diagnosis of cervical cancer. This held for women invited for screening, referred for colposcopy, and treated for a CIN lesion. A study on the impact of abnormal cervical smear results found that at baseline, shortly after the news, women who had received abnormal results reported significantly worse overall quality of life (EQ-5D and SF-6D) than women with normal results. At 12 weeks follow-up, only SF-6D results still significantly differed between groups. The QALYs lost during the 16 weeks after being informed of an abnormal smear result were estimated to be 0.007–0.009, which is equivalent to 2.4 to 3.2 days of healthy life lost.[42] Studies that focused on condition-specific effects like anxiety and worry did report negative effects of screen-detected non-cancer abnormalities.[6, 12, 42-44] Our findings may thus be related to the use of generic measures, which are less sensitive to small, condition-specific changes in health or quality of life. The patterns of the two generic measures used (EQ-5D and SF-6D) were roughly similar, although EQ-5D utility scores were often significantly higher than SF-6D scores. This is in line with the literature.[45] With different items and scoring mechanisms, the EQ-5D and the SF-6D were not expected to generate completely similar utilities. However, the EQ-5D and the SF-6D each resulted in a different preferred screening strategy, which clearly shows the dependence of CEA results on the choice of generic measure.

Limitations and strengths

With regard to the questionnaire study, we acknowledge that some of the (sub) groups were of limited size, that response rates are unavailable for the data collection among women referred for colposcopy and women diagnosed with cervical cancer, and that data considering the final stages of life were collected among patients with primary breast or lung cancer, and not among patients with cervical cancer. Also, we reported SF-6D as based on responses to the SF-12 items although SF-6D is better able to discriminate between conditions if based on responses to the SF-36 items.[46] Furthermore, screening has some rare undesired effects, like pregnancy complications after treatment for CIN, which were not captured in this study, but are important to be mentioned.[47] Also, we acknowledge that we assessed utilities in the Netherlands, and that we cannot determine if and to which extent this has influenced our results. However, we used widely accepted measures, that are each available in over 170 languages and that have been applied in numerous studies. We conclude that we have not solved all issues of how to determine the impact of cervical cancer on quality of life. However, at least we now have an idea about extent of the implications of screening and treatment for quality of life. We provided a set of empirical data and we do think that such a set is vital for the validity of CEAs. This data was collected in one country and we recommend more international data-collection to study external validity. Since we do not know what the results will be in another setting, we recommend to use the utilities as collected in the current study in modelling exercises, at least as an extra dataset, to enable sensitivity analyses. Still, this is the first study that made the extensive effort of comprising all health states induced or prevented by screening on cervical cancer. Data were collected in 7 cohorts, of which three were followed longitudinally. An additional strength is that the available clinical information enabled subdividing respondent groups according to FIGO stage. The design of our study has proven its feasibility and value, and can serve as a basis for measurements in other populations and situations.

This is also the first study focusing on the impact of utility assumptions on the recommended cervical cancer screening strategy. We, however, only used one model (for

the Dutch situation) to estimate the impact. Also, in absence of utility measurements for being HPV positive, we had to make the assumption that the effect on quality of life is similar for having a positive cytology test compared to having a positive HPV test. If higher quality of life losses will be found for being HPV positive than cytology positive, HPV screening might be less cost-effective than indicated by our calculations. Furthermore, in case of the HPV test, we assumed similar sensitivity for all disease stages with an HPV infection. Recent studies indicated lower sensitivity of HPV screening in women without any neoplasia.[48] We, however, calibrated our model to the observed HPV positivity rate (measured with a clinical validated HPV test) in the Dutch population, which determines the effect of HPV screening in the population. If we assume a lower sensitivity, we would have to increase the number of infections in the model, to reproduce that HPV positivity rate. Finally, the literature included in our analyses is not an exhaustive summary. We, for example, did not include HPV vaccination CEAs, even if vaccination was applied in a screening setting. Since the information on utility assumptions and the durations of the health states was sometimes lacking in the publications, we had to do some approximations to calculate the QADL. But even if our assumptions differ from the assumptions in the specific publications, our study clearly shows the importance of the utility assumptions for preferred cervical cancer screening strategies. We used the example of cervical cancer screening, but think the central message of our findings is generalizable to other settings, such as other (cancer) screening programs and vaccination programs.

Implications

The ongoing changes in cervical cancer prevention, such as the introduction of HPV screening and HPV vaccination programs require new policy, which is nowadays unacceptable without CEAs. The empirical utility scores that now have become available will enable more valid estimates of cost-effectiveness of screening programs and will contribute to a better evidence base for policy recommendations, which is especially useful given limited resources. Our results also showed that measuring utility values for different health states in screening is feasible but challenging, because these often involve small, condition-

specific changes in quality of life. This indicates that in economic evaluations, besides the QALY, other measurements that include the harms and benefits of screening (such as, number needed to screen or number needed to treat) need to be regarded as well.

Conclusion

This is the first study that empirically estimated utilities for the majority of the cervical cancer health states. We found that QoL assumptions in decision analyses vary in a range that is crucial for preferences regarding screening strategies. Empirical data show that primary HPV screening might not be the most cost-effective screening strategy, if QALYs are used in this decision. However, the empirical data also show that the measurement of utilities is challenging. This indicates that other measurements that include the harms and benefits of screening needs to be regarded as well.

Acknowledgements

We are grateful to the women who completed questionnaires for participating in the study. This work was funded by the Dutch Cancer Society (EMCR 2003–2775).

Role of the funding source

The authors were completely independent from funders in conducting and writing this study.

Conflict of interest statement

The authors have no conflicts of interest to disclose.

REFERENCES

1. IARC. Cervic cancer screening *IARC Handbook of cancer prevention*ed.: IARC Press, 2004.

2. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease] Principios y metodos del examen colectivo para identificar enfermedades. *Bol Oficina Sanit Panam* 1968;**65**: 281-393.

3. Gold M, Siegel JE, Russell L, Weinstein MC. *Cost-Effectiveness in Health and Medicine* ed. Oxford: Oxford University Press, 1996.

4. de Kok IM, van Rosmalen J, Dillner J, Arbyn M, Sasieni P, Iftner T, van Ballegooijen M. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. *BMJ* 2012;**344**: e670.

5. Naber SK, Matthijsse SM, Jansen EEL, de Kok IMCM, de Koning HJ, van Ballegooijen M. Effecten en kosten van het vernieuwde bevolkingsonderzoek naar baarmoederhalskanker in Nederland naar aanleiding van recente ontwikkelingen Rotterdam, 2016.

6. Korfage IJ, van Ballegooijen M, Huveneers H, Essink-Bot ML. Anxiety and borderline PAP smear results. *Eur J Cancer* 2010;**46**: 134-41.

7. Tahseen S, Reid PC. Psychological distress associated with colposcopy: patients' perception. *Eur J Obstet Gynecol Reprod Biol* 2008;**139**: 90-4.

8. Valdini A, Esielionis P. Measurement of colposcopy-associated distress using the psychosocial effects of having an abnormal pap smear-questionnaire in a Latina population. *J Low Genit Tract Dis* 2004;**8**: 25-32.

9. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD, Force I-SMGRPT. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. *Value Health* 2012;**15**: 835-42.

10. Korfage I, van Ballegooijen M, Wauben B, Looman C, Habbema J, Essink-Bot ML. Having a Pap smear, quality of life before and after cervical screening: a questionnaire study. *BJOG* 2012;**119**: 936-44.

11. Korfage IJ, Essink-Bot ML, Mols F, van de Poll-Franse L, Kruitwagen R, van Ballegooijen M. Health-related quality of life in cervical cancer survivors: a population-based survey. *Int J Radiat Oncol Biol Phys* 2009;**73**: 1501-9.

12. Korfage IJ, Essink-Bot ML, Westenberg SM, Helmerhorst T, Habbema JD, van Ballegooijen M. How distressing is referral to colposcopy in cervical cancer screening?: a prospective quality of life study. *Gynecol Oncol* 2014;**132**: 142-8.

13. Accetta G, Biggeri A, Carreras G, Lippi G, Carozzi FM, Confortini M, Zappa M, Paci E. Is human papillomavirus screening preferable to current policies in vaccinated and unvaccinated women? A cost-effectiveness analysis. *J Med Screen* 2010;**17**: 181-9.

14. Balasubramanian A, Kulasingam SL, Baer A, Hughes JP, Myers ER, Mao C, Kiviat NB, Koutsky LA. Accuracy and cost-effectiveness of cervical cancer screening by high-risk human papillomavirus DNA testing of self-collected vaginal samples. *J Low Genit Tract Dis* 2010;**14**: 185-95.

15. Berkhof J, Coupe VM, Bogaards JA, van Kemenade FJ, Helmerhorst TJ, Snijders PJ, Meijer CJ. The health and economic effects of HPV DNA screening in The Netherlands. *Int J Cancer* 2010;**127**: 2147-58.

16. Chuck A. Cost-effectiveness of 21 alternative cervical cancer screening strategies. *Value Health* 2010;**13**: 169-79.

17. Coupe VM, Bogaards JA, Meijer CJ, Berkhof J. Impact of vaccine protection against multiple HPV types on the cost-effectiveness of cervical screening. *Vaccine* 2012;**30**: 1813-22.

18. Coupe VM, de Melker HE, Snijders PJ, Meijer CJ, Berkhof J. How to screen for cervical cancer after HPV16/18 vaccination in The Netherlands. *Vaccine* 2009;**27**: 5111-9.

19. de Bekker-Grob EW, de Kok IM, Bulten J, van Rosmalen J, Vedder JE, Arbyn M, Klinkhamer PJ, Siebers AG, van Ballegooijen M. Liquid-based cervical cytology using ThinPrep technology: weighing the pros and cons in a cost-effectiveness analysis. *Cancer Causes Control* 2012;**23**: 1323-31.

20. Elbasha EH, Dasbach EJ, Insinga RP. A multi-type HPV transmission model. *Bull Math Biol* 2008;**70**: 2126-76.

21. Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. *J Natl Cancer Inst* 2008;**100**: 308-20.

22. Kitchener HC, Canfell K, Gilham C, Sargent A, Roberts C, Desai M, Peto J. The clinical effectiveness and cost-effectiveness of primary human papillomavirus cervical screening in England: extended follow-up of the ARTISTIC randomised trial cohort through three screening rounds. *Health Technol Assess* 2014;**18**: 1-196.

23. Huh WK, Williams E, Huang J, Bramley T, Poulios N. Cost effectiveness of human papillomavirus-16/18 genotyping in cervical cancer screening. *Applied Health Economics & Health Policy* 2015;**13**: 95-107.

24. Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N. Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis. *Health Technol Assess* 2004;**8**: iii, 1-78.

25. Kim JJ, Campos NG, Sy S, Burger EA, Cuzick J, Castle PE, Hunt WC, Waxman A, Wheeler CM, New Mexico HPVPRSC. Inefficiencies and High-Value Improvements in U.S. Cervical Cancer Screening Practice: A Cost-Effectiveness Analysis. *Ann Intern Med* 2015;**163**: 589-97.

26. Kitchener HC, Blanks R, Cubie H, Desai M, Dunn G, Legood R, Gray A, Sadique Z, Moss S, Group MTS. MAVARIC - a comparison of automation-assisted and manual cervical screening: a randomised controlled trial. *Health Technol Assess* 2011;**15**: iii-iv, ix-xi, 1-170.

27. Kulasingam SL, Myers ER, Lawson HW, McConnell KJ, Kerlikowske K, Melnikow J, Washington AE, Sawaya GF. Cost-effectiveness of extending cervical cancer screening intervals among women with prior normal pap tests. *Obstet Gynecol* 2006;**107**: 321-8.

28. Scoggins JF, Ramsey SD, Jackson JC, Taylor VM. Cost effectiveness of a program to promote screening for cervical cancer in the Vietnamese-American population. *Asian Pac J Cancer Prev* 2010;**11**: 717-22.

29. van Rosmalen J, de Kok IM, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. *BJOG* 2012;**119**: 699-709.

30. Voko Z, Nagyjanosi L, Margitai B, Kovi R, Toth Z, Laszlo D, Kalo Z. Modeling cost-effectiveness of cervical cancer screening in Hungary. *Value Health* 2012;**15**: 39-45.

31. van den Hout WB, Kramer GW, Noordijk EM, Leer JW. Cost-utility analysis of short- versus long-course palliative radiotherapy in patients with non-small-cell lung cancer. *J Natl Cancer Inst* 2006;**98**: 1786-94.

32. van den Hout WB, van der Linden YM, Steenland E, Wiggenraad RG, Kievit J, de Haes H, Leer JW. Single- versus multiple-fraction radiotherapy in patients with painful bone metastases: cost-utility analysis based on a randomized trial. *J Natl Cancer Inst* 2003;**95**: 222-9.

33. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;**42**: 851-59.

34. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;**35**: 1095-108.

35. Ronco G, Dillner J, Elfstrom KM, Tunesi S, Snijders PJ, Arbyn M, Kitchener H, Segnan N, Gilham C, Giorgi-Rossi P, Berkhof J, Peto J, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet* 2014;**383**: 524-32.

36. Simms K, Laprise J, Smith M, Lew J, Caruana M, Brisson M, Canfell K. Cost-effectiveness of the next generation nonavalent human papillomavirus vaccine in the context of primary human papillomavirus screening in Australia: a comparative modelling analysis. *The Lancet Public Health* 2016;**1**: e66 - e75

37. Lew JB, Simms K, Smith M, Lewis H, Neal H, Canfell K. Effectiveness Modelling and Economic Evaluation of Primary HPV Screening for Cervical Cancer Prevention in New Zealand. *PLoS One* 2016;**11**: e0151619.

38. Albrecht GL, Devlieger PJ. The disability paradox: high quality of life against all odds. *Social science & medicine* 1999;**48**: 977-88.

39. Tversky A, Kahneman D. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* 1991: 1039-61.

40. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of life Research* 2003;**12**: 599-607.

41. Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science* 1998;**9**: 340-6.

42. Drolet M, Brisson M, Maunsell E, Franco EL, Coutlee F, Ferenczy A, Fisher W, Mansi JA. The psychosocial impact of an abnormal cervical smear result. *Psychooncology* 2012;**21**: 1071-81.

43. Frederiksen ME, Njor S, Lynge E, Rebolj M. Psychological effects of diagnosis and treatment of cervical intraepithelial neoplasia: a systematic review. *Sexually transmitted infections* 2014: sextrans-2014-051754.

44. O'Connor M, Gallagher P, Waller J, Martin CM, O'Leary JJ, Sharp L. Adverse psychological outcomes following colposcopy and related procedures: a systematic review. *BJOG: An International Journal of Obstetrics & Gynaecology* 2016;**123**: 24-38.

45. Whitehurst DG, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Med Decis Making* 2011;**31**: E34-44.

46. Luo N, Wang P, Fu AZ, Johnson JA, Coons SJ. Preference-based SF-6D Scores Derived From the SF-36 and SF-12 Have Different Discriminative Power in a Population Health Survey. *Med Care* 2012;**50**: 627-32.

47. Habbema D, Weinmann S, Arbyn M, Kamineni A, Williams AE, I MCMdK, van Kemenade F, Field TS, van Rosmalen J, Brown ML. Harms of cervical cancer screening in the United States and the Netherlands. *Int J Cancer* 2017;**140**: 1215-22.

48. Rebolj M, Bonde J, Preisler S, Ejegod D, Rygaard C, Lynge E. Human Papillomavirus Assays and Cytology in Primary Cervical Screening of Women Aged 30 Years and Above. *PLoS One* 2016;**11**: e0147326.

Table 1. Number of quality adjusted days lost (QADL) (utility loss x duration in days) found in published CEAs and empirical data per health state. Publications with unknown durations of health states are excluded. [20, 28, 30]

Source	Induced by screening: screening and pre-invasive health states									Prevented by screening: invasive cancer health states			
	Prim. screens	Pos. prim. screens	False-pos. prim. screens	Triage tests	Referrals for colposcopy	False-pos. Referral	CIN1	CIN2	CIN3	FIGO 1	FIGO 2+	Palliative phase	Survivors
Empirical data EQ-5D	-	-	-	-	-	-	-	-	-	11.0	21.0	182.5	146.0
Empirical data SF-6D	-	-	-	13.7	0.9	-	-	-	-	29.2	51.1	182.5	109.5
Accetta[13]	-	-	-	-	-	-	-	-	-	116.8	167.3	-	-
Balasubramanian[14]	-	-	1.1	-	-	-	-	-	-	438.0	602.3	-	-
Berkhof[15]	-	0.9	-	-	-	-	5.5	25.6	25.6	113.2	328.5	-	-
Chuck[16]	-	-	-	-	-	-	32.9	47.5	47.5	178.9	232.2	-	-
Coupé[17, 18]	-	0.9	-	-	-	-	5.5	25.6	25.6	113.2	328.5	-	-
de Bekker-Grob[19]	0.1	-	-	0.1	-	0.9	5.5	25.6	25.6	113.2	511.0	21.9	-
de Kok[4]. van Rosmalen[29]	0.1	-	-	1.1	-	0.9	5.5	25.6	25.6	113.2	511.0	21.9	-
Goldhaber-Fiebert[21]	-	-	-	-	-	-	-	-	-	116.8	167.3	-	-
Karnon[24]	-	-	-	9.0	18.0	-	-	-	-	1460.0	1460.0	-	-
Kitchener[26]	-	-	14.6	-	-	-	40.2	43.8	40.2	438.0	767.5	-	-
Kitchener (Simonella)[22]	1.2	9.7	-	-	-	10.1	10.1	10.8	10.8	438.0	767.5	-	-
Kulasingam[27]	-	-	0.9	-	-	-	-	-	-	273.8	821.3	-	-

Accetta, Chuck and Goldhaber-Fiebert: Duration health states unknown, assumed 1 year; Chuck: Assumed per invasive stage: 1 year without treatment and 1 year with treatment, total duration 2 years; de Kok: Assumed 'duration since last test' in case of triage testis 6 months; Karnon: Based on 'This mortality is based on an average life expectancy with invasive cancer present in an unscreened population of approximately 10 years', we assumed a duration of 10 years for 'remainder of lifetime'; Kitchener: Similar durations assumed for cancer as in 'Kitchener_Simonella'; Kulasingam: Only QoL loss due to cancer treatment (duration 5 years) included.

Table 2. Quality of life based utility scores, obtained with SF-6D and EQ-5D, Dutch questionnaire study.

Stage	Assessments	SF-6D		EQ-5D		Paired t-test
		(0-1, SD)	CI (mean ±1.96SE)	(0-1, SD)	CI (mean ±1.96SE)	*p*-value
Reference						
1. General population	1. Between screening rounds	0.83 (0.13)	0.82-0.84	0.86 (0.20)	0.85-0.88	<0.001
Subgroup: screening ages (30-60 years)	1. Between screening rounds	0.84 (0.12)	0.83-0.85	0.88 (0.19)	0.86-0.89	<0.001
Screening						
2. Invited for screening	1. At invitation	0.84 (0.10)	0.84-0.85	0.89 (0.19)	0.88-0.90	<0.001
Subgroup: Screen participants	1. At invitation	0.84 (0.10)	0.84-0.85	0.89 (0.19)	0.88-0.91	<0.001
	2. After Pap test	0.85 (0.10)	0.85-0.86	0.90 (0.18)	0.89-0.91	<0.001
	3. After Pap test result	0.86 (0.11)	0.85-0.87	0.91 (0.17)	0.90-0.92	<0.001
Triage						
3. Receiving a repeat Pap test after BMD[1.] result	1. 6-24 months after BMD[1.] Pap test	0.80 (0.13)	0.79-0.82	0.87 (0.21)	0.84-0.89	<0.001
Subgroup: NOT (yet) referred for colposcopy	1. 6-24 months after BMD Pap test	0.81 (0.12)	0.79-0.83	0.88 (0.21)	0.84-0.91	<0.001
Referral for colposcopy						
4. Initial six months following referral for colposcopy	1. Shortly after suspicious Pap test	0.81 (0.10)	0.79-0.83	0.91 (0.13)	0.89-0.94	<0.001
	2. At 1 month f-up	0.82 (0.12)	0.79-0.84	0.89 (0.18)	0.86-0.93	<0.001
	3. At 3 months f-up	0.82 (0.10)	0.80-0.85	0.92 (0.17)	0.89-0.95	<0.001
	4. At 6 months f-up	0.83 (0.12)	0.81-0.86	0.91 (0.20)	0.87-0.95	<0.001
5. 6-35 months following referral for colposcopy	1. 6-35 months after treatment of CIN	0.83 (0.11)	0.80-0.85	0.91 (0.15)	0.88-0.94	<0.001

(Follow up) Table 2. Quality of life based utility scores, obtained with SF-6D and EQ-5D, Dutch questionnaire study.

Cervical cancer						
6. Diagnosed with cervical cancer: FIGO[2.] 1A&1B	1. After diagnosis	0.73 (0.13)	0.69-0.77	0.79 (0.21)	0.73-0.85	0.004
	2. At 3 months f-up	0.75 (0.14)	0.70-0.79	0.81 (0.14)	0.77-0.85	0.001
	3. At 6 months f-up	0.76 (0.15)	0.71-0.81	0.82 (0.21)	0.76-0.89	0.031
	4. At 12 months f-up	0.81 (0.13)	0.77-0.85	0.87 (0.13)	0.83-0.91	0.016
Diagnosed with cervical cancer: FIGO[2.] 2+	1. After diagnosis	0.71 (0.18)	0.62-0.80	0.72 (0.34)	0.55-0.89	0.939
	2. At 3 months f-up	0.63 (0.11)	0.57-0.70	0.63 (0.31)	0.46-0.80	0.628
	3. At 6 months f-up	0.71 (0.13)	0.64-0.79	0.74 (0.24)	0.61-0.87	0.621
	4. At 12 months f-up	0.75 (0.19)	0.63-0.87	0.73 (0.38)	0.51-0.94	0.603
7. Tumorfree, 2-10 years after diagnosis	1. Disease free	0.80 (0.14)	0.79-0.82	0.82 (0.25)	0.79-0.85	0.007
8. Having advanced cancer	1. Last year of life Breast cancer Lung cancer 2. Last three months of life Breast cancer Lung cancer	-	-	 0.36 (0.37) 0.11 (0.38) 0.13 (0.39) 0.10 (0.36)	 0.33-0.40 0.02-0.21 0.08-0.18 -0.01-0.21	-

[1.]BMD=borderline or mild dyskaryosis;[2.]FIGO= International Federation of Gynecology and Obstetrics; f-up relates to follow-up in terms of timing of questionnaire assessments

Table 3. Preferred screening strategy at a threshold of €20,000 and €50,000 per QALY gained (3% discounting) by different sources (i.e. literature and empirical data) of QoL assumptions. Costs and QALYs gained presented are absolute, compared to no screening.

Threshold	QoL assumptions based on	Primary test	Start age	Number of tests	Interval (years)	Costs (€, x1000)	QALYs gained	ICER (€)
€20,000 per QALY gained	Empirical data EQ-5D	HPV	30	4	6	369	63	17,857
	Empirical data SF-6D	Cytology	32	5	5	409	63	19,963
	Accetta[13]	HPV	30	4	6	369	62	18,007
	Balasubramanian[14]	HPV	30	4	6	369	70	16,350
	Berkhof[15]	HPV	30	4	6	369	60	18,901
	Chuck[16]	HPV	30	4	6	369	57	20,420
	Coupe[17, 18]	HPV	30	4	6	369	60	18,901
	de Bekker-Grob[19]	HPV	30	4	6	369	62	18,844
	de Kok[4], van Rosmalen[29]	HPV	30	4	6	369	62	19,220
	Goldhaber-Fiebert[21]	HPV	30	4	6	369	62	18,007
	Karnon[24]	Cytology	30	6	4	472	89	17,151
	Kitchener[26]	Cytology	32	4	5	288	56	16,099
	Kitchener (SImonella)[22]	HPV	32	3	6	252	42	16,100
	Kulasingam[27]	HPV	30	5	6	520	79	19,706
€50,000 per QALY gained	Empirical data EQ-5D	HPV	27	9	5	1,074	84	47,378
	Empirical data SF-6D	Cytology	27	10	4	931	79	47,739
	Accetta[13]	HPV	27	9	5	1,074	83	47,977
	Balasubramanian[14]	HPV	27	9	5	1,074	92	41,426
	Berkhof[15]	HPV	30	7	6	833	75	43,015
	Chuck[16]	HPV	30	7	6	833	70	48,966
	Coupe[17, 18]	HPV	30	7	6	833	75	43,015
	de Bekker-Grob[19]	HPV	27	8	5	921	78	49,879
	de Kok[4], van Rosmalen[29]	HPV	30	7	6	833	75	46,202
	Goldhaber-Fiebert[21]	HPV	27	9	5	1,074	83	47,977
	Karnon[24]	Cytology	30	10	4	992	107	41,144
	Kitchener[26]	Cytology	27	7	5	586	66	39,120
	Kitchener (SImonella)[22]	HPV	32	3	6	252	42	16,100
	Kulasingam[27]	HPV	27	9	5	1,074	95	40,037

Figure legends

Figure 1. Utilities measured by the SF-6D and EQ-5D questionnaire per stage in the screening and follow up process. Dutch questionnaire study.

Figure 2. Net costs and health effects (compared to no screening) of the efficient screening programmes for the fourteen different QADL-sets, for a cohort of 100,000 unvaccinated women, for the period from 2011 onwards. Costs and effects discounted with 3% towards year 2011. Each point corresponds with a screening policy on the efficient frontier. Black marker points represent the policy at the €20,000 (left-side on the frontier) and €50,000 (right-side at the frontier) per quality-adjusted life year (QALY) gained threshold (see table 3).

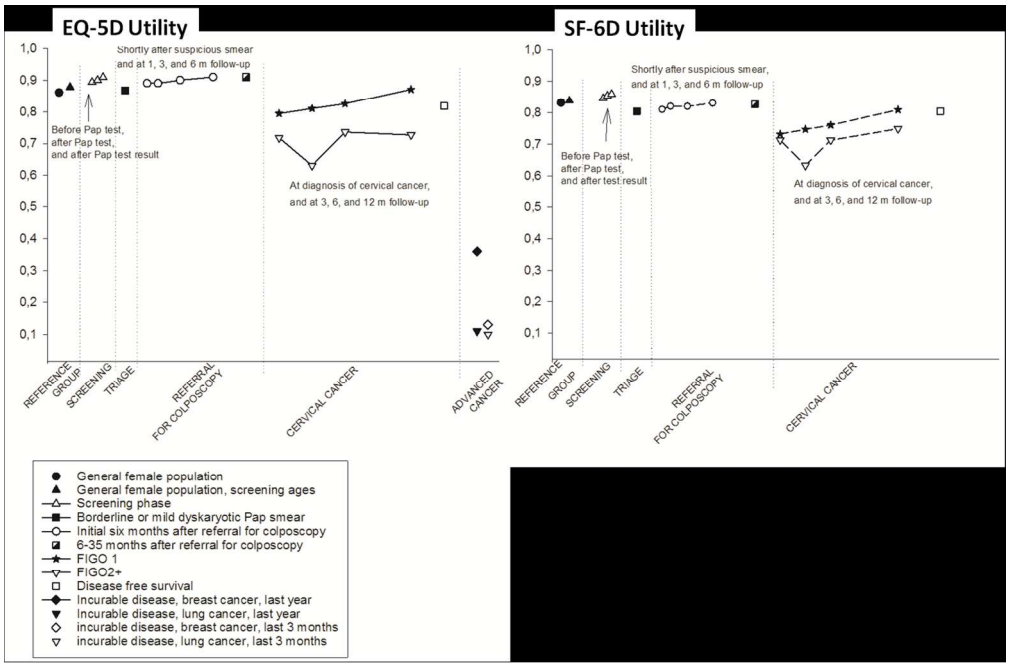
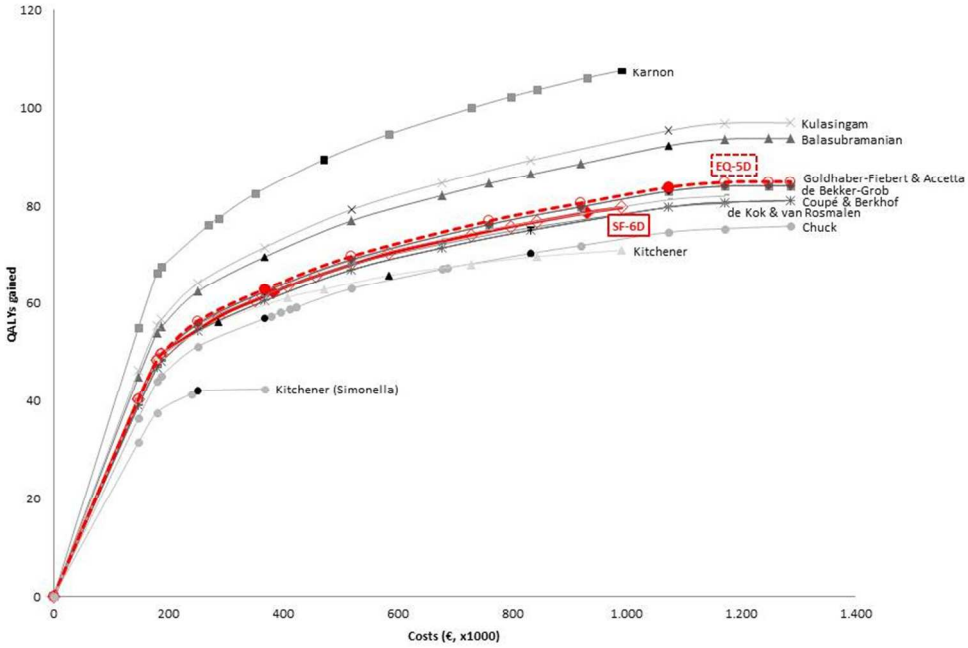


Figure 1

271x178mm (150 x 150 DPI)



Net costs and health effects (compared to no screening) of the efficient screening programmes for the fourteen different QADL-sets, for a cohort of 100,000 unvaccinated women, for the period from 2011 onwards. Costs and effects discounted with 3% towards year 2011. Each point corresponds with a screening policy on the efficient frontier. Black marker points represent the policy at the €20,000 (left-side on the frontier) and €50,000 (right-side at the frontier) per quality-adjusted life year (QALY) gained threshold (see table 3).

254x190mm (96 x 96 DPI)