

COLORECTAL CANCER SCREENING

VOLUME 17

This publication represents the views and expert opinions of an IARC Working Group on the Evaluation of Cancer-Preventive Strategies, which met in Lyon, 14–21 November 2017

LYON, FRANCE - 2019

IARC HANDBOOKS OF
CANCER PREVENTION

3. STUDIES OF COLORECTAL CANCER SCREENING

3.1 Methodological considerations

Methods for colorectal cancer (CRC) screening include endoscopic methods and stool-based tests for blood. The two primary end-points for endoscopic CRC screening are (i) finding cancer at an early stage (secondary prevention) and (ii) finding and removing precancerous lesions (adenomatous polyps), to reduce the incidence of CRC (primary prevention). The primary end-point for stool-based tests is finding cancer at an early stage. Stool-based tests also have some ability to detect adenomatous polyps; therefore, a secondary end-point of these tests is reducing the incidence of CRC.

3.1.1 *Randomized controlled trials of colorectal cancer screening*

A randomized controlled trial (RCT) that compares a screening arm with a non-screening arm (or with a screening arm with a different screening modality) is considered the reference standard in evaluating the cancer-preventive effects of a screening test. For screening modalities that only detect cancerous lesions (such as those for breast cancer or lung cancer), the primary end-point of a screening RCT is generally mortality from the cancer of interest. For screening modalities that have the potential to detect and remove cancer precursors (such as those for CRC and cervical cancer), a co-primary

end-point of the RCT can be the incidence of the cancer of interest.

The observed effect of screening in RCTs is dependent on, among other things, the participation in the intervention group and the limitation of contamination of the control group. Low participation biases the estimate of effect towards its no-effect value, and therefore it must be evaluated and reported. Screening of controls by services outside of the RCT also dilutes the effect of screening on CRC incidence and/or mortality. If the screening modality being evaluated is widely used in clinical practice in the region or regions where the RCT is being conducted, then contamination may be considerable, although it may be difficult and/or costly to estimate its extent. The standard evaluation of the primary end-point is an intention-to-treat analysis. Methods to adjust for contamination and low participation, called per-protocol analyses, have been proposed ([Cuzick et al., 1997](#); [Baker et al., 2002](#)). [Note that these methods differ from those that compare participants who were actually screened with those who were not, which is an invalid method with a high potential for selection bias.]

3.1.2 *Observational studies on the preventive effects of colorectal cancer screening*

Data from observational studies, in which study participants are not randomly allocated to be screened or not screened, should be used with caution, because of their greater propensity to bias. Comparison of the survival, or stage distribution, of screen-detected cases with that of cases diagnosed outside of screening is notoriously flawed, because of lead-time bias, length-time bias, and overdiagnosis bias. Lead time is the period between when a cancer is found by screening and when it would have been detected from clinical signs and symptoms (not directly observable) in the absence of screening. Survival time, by definition, is measured from the date of diagnosis to the date of death. Lead-time bias is the overestimation of survival time due to earlier detection by screening than clinical presentation; it is very difficult to distinguish from earlier diagnosis leading to a real extension of life. Length-time bias reflects disproportionate detection by screening of indolent tumours, which may reside in a preclinical state for longer time periods than aggressive tumours do. Overdiagnosis is an extreme form of length-time bias in which a tumour that would never have been diagnosed without screening is detected by screening.

Using cohort or case-control designs to compare the mortality and/or incidence rates of a group receiving (or invited to receive) screening with those of a group not receiving (or not invited to receive) screening can avoid the above-mentioned biases, but it will usually involve some type of selection bias, because the decision to be screened is not random and may be related to factors that predispose to, or against, developing the cancer. Careful adjustment for known CRC risk factors may partially alleviate this bias. Cohort studies are generally more reliable than case-control studies for assessing the efficacy of screening, although well-designed case-control studies may be useful. Comparison of the

incidence or mortality rates in a screened cohort with concurrent (or past) population rates is also prone to the selection biases described above and may also be biased by variation in temporal and regional trends.

Ecological studies that compare incidence or mortality rates in a region or country where screening has been implemented with rates in the same region or country during a previous time period, or with contemporaneous rates in a neighbouring region or country, may be useful but are subject to the standard caveats of ecological studies.

There are several early observational studies that assessed the effects of endoscopy and stool-based tests for blood on CRC incidence and mortality; most of these studies have major methodological issues. Lead-time bias, length-time bias, selection bias, and confounding factors in these studies could lead to overestimation or underestimation of the effect when not adjusted for. Misclassification of the outcome, CRC incidence or death (ascertainment bias), can also bias the effect estimates. In the Effectiveness of Screening for Colorectal Cancer in Average-Risk Adults (SCOLAR) nested case-control study, [Goodman et al. \(2015\)](#) stressed the challenges in observational studies of colonoscopy screening, including distinguishing between indications for having had a colonoscopy (screening or investigation of symptoms). Some observational studies use, for example, a 6-month window to exclude index examinations that were done to investigate symptoms of CRC.

In light of the published RCTs of sigmoidoscopy screening, the Working Group established two criteria for observational studies to be included in the evaluation of effectiveness: the study (i) must be performed in a screening setting and (ii) must not exclude cancer detected at the baseline endoscopy. Moreover, the Working Group established six considerations to weigh the impact of individual studies on the overall estimate: (i) there must be a concurrent control

group, (ii) there must be an adequate length of follow-up, (iii) the sample size must be large enough to detect relevant effects, (iv) the study must be conducted in a setting with contemporary methods, (v) the outcome (and exposure) ascertainment must be reliable, and (vi) potential confounder data must be available and adjusted for in the analysis.

3.1.3 Evaluation of the adverse effects of colorectal cancer screening

An adverse effect of screening is defined as any negative effect on individuals or populations that results from being involved in the screening process compared with not being involved in screening. It is important to quantify not only the frequency of the harm but also its magnitude. It is also important to recognize that the harms in an RCT may be different from the harms in a screening programme.

(a) Definitions of harms

[Harris et al. \(2014\)](#) proposed a taxonomy of the harms of screening. They proposed four domains of harms: physical effects, psychological effects, financial strain, and opportunity costs. Financial strain and opportunity costs [the indirect effect of screening on health-related activities] are not considered in this review. Here, harms are classified on the basis of where in the screening cascade they occur: harms associated with (i) the screening process, (ii) the screening test itself, and (iii) the management of a positive screening result.

Potential harms of the screening process include anxiety, caused by the invitation to screening or awaiting the results of screening, and a negative impact on lifestyle or health-seeking behaviour. The assessment of these outcomes requires validated instruments and is best conducted as a longitudinal study over the course of screening, ideally as part of an RCT,

comparing individuals invited to screening with those not invited to screening.

It is possible that reassurance from a negative screening result (whether true-negative or false-negative) could lead to delay in presentation for investigation of symptoms that develop in the interval between scheduled screening tests, with consequent late diagnosis of an interval cancer and possibly death from it. Such an impact on mortality would form part of the overall mortality measured in the screening arm of an RCT of CRC screening and would not be measurable separately from it. However, its presence within the results of the RCT would remove any need to account for it separately when estimating a benefit-harm ratio from the RCT.

No physical harms are incurred with stool-based tests for blood. Potential harms of endoscopic tests include pain, physical damage to the bowel due to endoscopy, possible hospitalization, and the need for surgical repair. The frequency and severity of such harms can be estimated from a representative screened cohort.

Potential harms associated with the management of a positive screening result include physical harm from the workup and treatment and harm from the psychological response to knowledge of the result and any aspect of its management. Apart from the possible harms of treating a screen-detected cancer, the harms experienced are largely independent of whether the screening result is true-positive or false-positive. Some people restrict the definition of a true-positive to invasive cancer, whereas others include advanced adenoma in the definition, and still others may include detection and removal of precancerous lesions (polypectomy).

(b) Overdiagnosis

An overdiagnosed cancer is defined as a cancer detected as a result of screening that would not have been diagnosed if screening had not taken place. The lower the pathological grade of a screen-detected cancer and the shorter

the individual's life expectancy at the time the cancer is diagnosed, the more likely it is that the cancer is an overdiagnosed cancer. The harm associated with overdiagnosis comes from labelling the individual as a cancer patient and from any adverse effects of the treatment of the cancer.

(i) Quantification of overdiagnosis

Screening that leads to the earlier diagnosis of cancer will always cause an apparent increase in cancer incidence, some of which is due to advancing the time of diagnosis of cancers that would have been diagnosed anyway, and some of which is due to overdiagnosis. To quantify overdiagnosis accurately, it would be necessary to follow up identical screened and unscreened cohorts until the rise in the cancer incidence rate in the screened cohort has stabilized, and then stop screening and continue to follow up the cohorts until the cancer incidence rates are approximately the same in the screened and unscreened cohorts. The total excess rate of cancer in the screened cohort is then the rate of cancer overdiagnosis. In practice, such a quantification process is rarely possible, although it could be approximated by a well-done RCT that adopts most of the above-mentioned features. There are various other ways to estimate overdiagnosis, but few or none that are likely to produce a highly accurate result ([Carter et al., 2015](#); [Ripping et al., 2017](#); [Davies et al., 2018](#)).

When quantifying overdiagnosis, one may consider it (i) as a cumulative lifetime risk of overdiagnosis associated with participation in screening; (ii) as the ratio of the cumulative rate of overdiagnosis to the cumulative rate of cancer in unscreened individuals, expressed as a percentage; (iii) as the percentage of cancers in screened individuals that are overdiagnosed; or (iv) as the percentage of screen-detected cancers that are overdiagnosed. In a screening programme that also *prevents* cancer (via detection and treatment of precursors), so that screened people have a lower cumulative rate of diagnosis

of cancer than unscreened people, it is practically impossible to determine whether some of the screen-detected cancers are overdiagnosed cancers. Although the reduction in cancer incidence indicates a clear benefit, some harm from the detection and treatment of precancerous lesions that would never have become symptomatic invasive cancers in the person's lifetime is also likely.

There are three types of harms of overdiagnosis as usually defined: (i) labelling of an individual as a cancer patient (which may cause anxiety and lead to problems obtaining health insurance and life insurance); (ii) the immediate side-effects of treatment; and (iii) the long-term consequences of diagnosis, including, for example, intensive surveillance. Quantification of these harms is an important contributor to obtaining an accurate estimate of the balance of benefits and harms of screening. Each of these types of harms will apply to some degree to the detection and treatment of simple polyps or adenomas at varying degrees of advancement during the course of screening for CRC.

(ii) Overdiagnosis of precancerous lesions

The proportion of polyps that would cause symptoms if not removed is low but increases as one moves from non-adenomatous polyps to non-advanced adenomas and to advanced adenomas (including multiple non-advanced adenomas). The Working Group could not agree about whether to use the term overdiagnosis to include simple polyps, non-advanced adenomas, or even adenomas. Some felt that it should be possible to estimate how many adenomas would not have progressed to cancer and that this number should be reported as being overdiagnosed. (That group did not wish to attempt to quantify how many adenomas may have been diagnosed in the absence of screening, as a result of colonoscopies in response to bowel symptoms.) Other Working Group members felt that the term overdiagnosis was not helpful and that

it was more appropriate to report the number of polyps, non-advanced adenomas, and advanced adenomas and the consequences of each of those diagnoses. It was agreed that the frequency of detection of polyps, non-advanced adenomas, and advanced adenomas should be reported, as should the consequences of such detection, including potential harms and their frequency.

There was a lack of agreement among Working Group members about whether some or all of the harms associated with treatment of these precancerous lesions should be attributed to screening.

3.1.4 Interval cancers

An interval cancer is defined as a cancer that is diagnosed between routine screens, that is, cancers diagnosed after a positive screening test that were not diagnosed as a result of that screening test. Thus, interval cancers are cancers that were missed by screening or by investigation after a positive screening test, or that have developed since the most recent screening test was done.

For most screening, the interval cancer rate will depend on the screening interval: the shorter the interval, the smaller the chance of developing an interval cancer. For this reason, it may be useful to consider cancers diagnosed x years after a negative screening result (where $x = 1, 2, 3, \dots$) and to censor the follow-up when the individual is next screened. The interval cancer rate may also be higher after the first screen than it is after subsequent screens, because of the higher prevalence of cancer at the time of the first screen and a consequently higher risk of an incident cancer in the first interval between screens.

Although interval cancers are considered by some people to be a harm of screening, they are better conceived of as a lack of benefit due to the (probably unavoidable) imperfect sensitivity of the test. There is no suggestion that an interval cancer would not also have been diagnosed in the absence of screening. Nonetheless, interval

cancers are an indicator of test insensitivity, and if the interval cancer rate in the screened population is greater than the quality assurance benchmark for the test, reasons for this performance deficiency should be sought and addressed.

3.1.5 Benefit–harm ratio and cost–effectiveness

Screening should be implemented only if its benefits outweigh its harms and if the financial resource requirements are reasonable in relation to the net benefit of screening (i.e. if screening is cost-effective).

(a) Benefit–harm ratio

All forms of screening are associated with both benefits and harms. The updated sets of criteria for screening published by the World Health Organization ([Andermann et al., 2008](#)) and the United States Preventive Services Task Force ([Harris et al., 2011](#)) both explicitly mention the balance between benefits and harms as a decisive criterion for the implementation of a screening intervention. The difficulty with this criterion is how to objectively weigh the benefits and harms of screening, because they are measured differently. How much harm is acceptable for every CRC death, or every incident CRC case, that can be prevented by screening? A growing number of studies consistently find that the attributes of screening tests, such as efficacy, process, and cost, are significant determinants of the choice of implementing a screening programme and of the screening method ([Mansfield et al., 2016](#)).

Informed decisions with respect to screening can be aided by giving people considering screening an outcomes table, which presents quantitative information about all potential benefits and harms of screening, so that people can decide whether they want to participate in screening. Such an outcomes table is a useful tool for patient information, but governments or guideline-issuing institutes still need to

determine for the population as a whole whether the benefit–harm ratio is favourable. This can be done simply, but subjectively, by committee consensus or, more objectively, by translating benefits and harms into summary measures of health benefit or harm, such as quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) gained or lost (see below). This has been done very rarely.

(b) *Cost–effectiveness*

QALYs and DALYs are two summary measures of health or life gained or lost that can be used to estimate the net outcome of health interventions. They each incorporate years of life and quality of the life in a single measure. QALYs are a measure, in the aggregate, of the gain or loss of life of a group subject to the intervention weighted by a measure of the quality of the life. Quality of life, for this purpose, is generally measured in terms of a person’s ability to carry out the activities of daily life, free of pain and mental disturbance, with a weight of 1 for fully able and a weight of 0 for totally unable. One DALY can be thought of as one *lost* year of disability-free life, where years of life lost to death have a weight of 1 and years of life lived free of any disability have a weight of 0, with the experience of a group measured in DALYs aggregated, as described above for QALYs.

To extrapolate outcomes from RCTs to lifetime QALYs and DALYs, mathematical models are typically used to track the benefits and harms of the intervention, accounting in the QALY weights used for any adverse effects of the intervention that intervention-arm participants experienced, taking account of the deaths, aggregating the QALYs or DALYs in each arm, and calculating the difference between the two arms as the net benefit of the intervention expressed as QALYs gained or DALYs averted, each of which can be expressed as positive or negative values.

Cost–effectiveness analysis formally compares the health outcomes with the economic

costs of different interventions, thereby assisting decision-makers to identify the interventions that will yield the greatest health benefits, given their resource constraints ([Cantor, 1994](#)). The costs that are considered depend on the perspective but generally include not only those of the intervention itself (in this case, screening) but also those of diagnostic follow-up and workup and adverse effects. Potential longer-term savings from treatment leading to prevented cases of disease (or advanced disease) are also included. The results of a cost–effectiveness analysis are summarized in an incremental cost–effectiveness ratio. The QALYs gained or DALYs averted with a particular strategy (in this *Handbook*, a CRC screening prevention programme) compared with an alternative strategy (no CRC screening programme) are included in the denominator, and the additional (incremental) costs of that strategy (compared with the same alternative) are included in the numerator, yielding an incremental cost per DALY averted or per QALY gained ([Cantor, 1994](#)). Future costs and benefits of the intervention are usually discounted to their present value ([Sanders et al., 2016](#)).

To ensure efficient use of resources, the cost–effectiveness of an intervention such as screening should be compared not only with the situation without screening but also, if applicable, with alternative screening interventions. For example, the costs and effects of endoscopic screening should be compared with those of stool-based screening strategies, and the costs and effects of 10-yearly endoscopy should be compared with those of single endoscopy. The incremental cost–effectiveness ratio is also used for this comparison.

The World Health Organization principles for population screening state that screening should be implemented only when there is a good balance between costs and benefits ([Wilson & Jungner, 1968](#)). Unfortunately, there is no universal definition of “good balance”; for example, in the USA an intervention that provides an additional year

of life at an incremental cost of US\$ 100 000 per life year gained has been accepted as a reasonable balance between costs and effects ([Weinstein, 2008](#)). The National Institute for Health and Care Excellence in the United Kingdom considers interventions with an incremental cost–effectiveness ratio of less than £20 000 per QALY gained to be cost-effective ([NICE, 2013](#)). Interventions with cost–effectiveness ratios in the range of £20 000 to £30 000 per QALY gained can still be considered acceptable, depending on additional criteria, whereas interventions costing more than £30 000 per QALY gained are generally not considered to be cost-effective. The World Health Organization has suggested using cost–effectiveness thresholds of 1–3 times the annual per capita gross domestic product of the country ([WHO, 2014](#)). Such an approach is particularly relevant in cost–effectiveness studies for CRC screening (see Sections 3.2.6, 3.3.6, and 3.5.4), which have been performed mostly in high-income countries. Consequently, the findings with respect to costs and benefits of screening may be very different in low- and middle-income countries, which generally have lower background cancer risks, different cost levels, and vastly different abilities to pay.

(c) *Using cost–effectiveness analyses to determine age limits and intervals for screening*

Cost–effectiveness analyses can address more than just the question of whether a certain intervention is cost-effective. RCTs are the reference standard for evaluation of the effectiveness of interventions; however, only a limited number of strategies can be evaluated in RCTs, whereas the number of different potential strategies is endless. Screening strategies can differ in the test modality used, the age at which to begin screening, the age at which to end screening, and the screening interval. Valid cost–effectiveness models offer the opportunity to extrapolate beyond the observations in the RCTs and assess

and compare alternative intervention strategies in an efficient way. Furthermore, such models can estimate the budget and resource impact of these strategies, so that only strategies that are feasible in a particular setting can be considered. For example, for the implementation of the CRC screening programme in the Netherlands, cost–effectiveness modelling had shown that the most cost-effective screening strategy to implement was faecal immunochemical test (FIT) screening with a low cut-off level for a positive test ([Wilschut et al., 2011a](#)). However, in case of limited colonoscopy capacity, the modelling indicated that the most cost-effective alternative would be to offer FIT screening with a higher cut-off level for a positive test ([Wilschut et al., 2011b](#)).

References

- Andermann A, Blancquaert I, Beauchamp S, Déry V (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*, 86(4):317–9. doi:[10.2471/BLT.07.050112](#) PMID:[18438522](#)
- Baker SG, Kramer BS, Prorok PC (2002). Statistical issues in randomized trials of cancer screening. *BMC Med Res Methodol*, 2(1):11. doi:[10.1186/1471-2288-2-11](#) PMID:[12238954](#)
- Cantor SB (1994). Cost-effectiveness analysis, extended dominance, and ethics: a quantitative assessment. *Med Decis Making*, 14(3):259–65. doi:[10.1177/0272989X9401400308](#) PMID:[7934713](#)
- Carter JL, Coletti RJ, Harris RP (2015). Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ*, 350:g7773. doi:[10.1136/bmj.g7773](#) PMID:[25569206](#)
- Cuzick J, Edwards R, Segnan N (1997). Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med*, 16(9):1017–29. doi:[10.1002/\(SICI\)1097-0258\(19970515\)16:9<1017::AID-SIM508>3.0.CO;2-V](#) PMID:[9160496](#)
- Davies L, Petitti DB, Martin L, Woo M, Lin JS (2018). Defining, estimating, and communicating overdiagnosis in cancer screening. *Ann Intern Med*, 169(1):36–43. doi:[10.7326/M18-0694](#) PMID:[29946705](#)
- Goodman M, Fletcher RH, Doria-Rose VP, Jensen CD, Zebrowski AM, Becerra TA, et al. (2015). Observational methods to assess the effectiveness of screening colonoscopy in reducing right colon cancer mortality risk:

- SCOLAR. *J Comp Eff Res*, 4(6):541–51. doi:[10.2217/cer.15.39](https://doi.org/10.2217/cer.15.39) PMID:[26201973](https://pubmed.ncbi.nlm.nih.gov/26201973/)
- Harris R, Sawaya GF, Moyer VA, Calonge N (2011). Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. *Epidemiol Rev*, 33(1):20–35. doi:[10.1093/epirev/mxr005](https://doi.org/10.1093/epirev/mxr005) PMID:[21666224](https://pubmed.ncbi.nlm.nih.gov/21666224/)
- Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, et al. (2014). The harms of screening: a proposed taxonomy and application to lung cancer screening. *JAMA Intern Med*, 174(2):281–5. doi:[10.1001/jamainternmed.2013.12745](https://doi.org/10.1001/jamainternmed.2013.12745) PMID:[24322781](https://pubmed.ncbi.nlm.nih.gov/24322781/)
- Mansfield C, Tangka FK, Ekwueme DU, Smith JL, Guy GP Jr, Li C, et al. (2016). Stated preference for cancer screening: a systematic review of the literature, 1990–2013. *Prev Chronic Dis*, 13:E27. doi:[10.5888/pcd13.150433](https://doi.org/10.5888/pcd13.150433) PMID:[26916898](https://pubmed.ncbi.nlm.nih.gov/26916898/)
- NICE (2013). Guide to the methods of technology appraisal 2013. Process and Methods Guides No. 9. London, UK: National Institute for Health and Care Excellence. Available from: <https://www.nice.org.uk/process/pmg9/>.
- Ripping TM, Ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broeders MJM (2017). Quantifying overdiagnosis in cancer screening: a systematic review to evaluate the methodology. *J Natl Cancer Inst*, 109(10):dxj060. doi:[10.1093/jnci/djx060](https://doi.org/10.1093/jnci/djx060) PMID:[29117353](https://pubmed.ncbi.nlm.nih.gov/29117353/)
- Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. (2016). Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA*, 316(10):1093–103. doi:[10.1001/jama.2016.12195](https://doi.org/10.1001/jama.2016.12195) PMID:[27623463](https://pubmed.ncbi.nlm.nih.gov/27623463/)
- Weinstein MC (2008). How much are Americans willing to pay for a quality-adjusted life year? *Med Care*, 46(4):343–5. doi:[10.1097/MLR.0b013e31816a7144](https://doi.org/10.1097/MLR.0b013e31816a7144) PMID:[18362811](https://pubmed.ncbi.nlm.nih.gov/18362811/)
- WHO (2014). Cost-effectiveness and strategic planning (WHO-CHOICE): threshold values for intervention cost-effectiveness by region. Geneva, Switzerland: World Health Organization. Available from: <https://www.who.int/choice/costs/en/>.
- Wilschut JA, Habbema JD, van Leerdam ME, Hol L, Lansdorp-Vogelaar I, Kuipers EJ, et al. (2011b). Fecal occult blood testing when colonoscopy capacity is limited. *J Natl Cancer Inst*, 103(23):1741–51. doi:[10.1093/jnci/djr385](https://doi.org/10.1093/jnci/djr385) PMID:[22076285](https://pubmed.ncbi.nlm.nih.gov/22076285/)
- Wilschut JA, Hol L, Dekker E, Jansen JB, Van Leerdam ME, Lansdorp-Vogelaar I, et al. (2011a). Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology*, 141(5):1648–55.e1. doi:[10.1053/j.gastro.2011.07.020](https://doi.org/10.1053/j.gastro.2011.07.020) PMID:[21784045](https://pubmed.ncbi.nlm.nih.gov/21784045/)
- Wilson JMG, Jungner G (1968). Principles and practice of screening for disease. Public Health Papers No. 34. Geneva, Switzerland: World Health Organization. Available from: http://whqlibdoc.who.int/php/WHO_PHP_34.pdf.