Supplemental Table 1: Modified PICOTS Criteria for Study Eligibility

	Include	Exclude
Population	Studies that attempt to measure, estimate, or quantify the amount of overdiagnosis resulting from a cancer screening test in an asymptomatic population Cancer types eligible for inclusion: prostate, breast, lung, colon, melanoma, bladder, renal, thyroid, uterine Studies that look at biologic characteristics of tumors (i.e., grade, doubling time) and draw conclusions about an amount of overdiagnosis	Studies addressing the potential for overdiagnosis but that do not draw conclusions regarding an amount of overdiagnosis (for example, studies that report on prevalence of early stage cancer detected at autopsy) Studies investigating different thresholds for tumor markers that comment on implications for overdiagnosis
Intervention	Method for measuring, estimating, or quantifying overdiagnosis	
Outcome	Magnitude of overdiagnosis	
Time Frame	Studies performed over any time frame	
Setting	Any setting	
Study Design	Randomized controlled trials, prospective or retrospective cohort studies, ecologic studies, case control studies, modeling studies Systematic reviews that identify other types of data (such as incidence trends) and use this to estimate overdiagnosis	Non-systematic reviews, case reports, case series Systematic reviews that only compile results from other studies that quantify overdiagnosis

Supplemental Table 2: Search Strategy

Database	Search Terms
PubMed	(cancer*[tw] OR neoplasms[MeSH]) AND (Screening*[tw] OR early diagnos*[tw] OR early detect*[tw]) AND (overdiagnos*[tw] OR over diagnos*[tw] OR overdetect*[tw] OR over detect*[tw])
Embase	(cancer*:ti,ab,de OR neoplasm*:ti,ab OR 'neoplasm'/exp) AND (screening*:ti,ab,de OR “early diagnosis”:ti,ab,de OR “early detection”:ti,ab,de) AND (overdiagnos*:ti,ab,de OR “over diagnosis”:ti,ab,de OR overdetect*:ti,ab,de OR “over detection”:ti,ab,de)

Supplemental Table 3: Standard Criteria for Evaluating Risk of Bias, by Study Design

Cohort and Ecologic Studies (adapted from Harris et al, 2011[7])

A. Risk of Bias *(rate overall as high/moderate/low)*
    i. Probability of selection bias and confounding *(rate as high/moderate/low)*
        i. Unbiased creation of comparable groups (at least after adjustment), especially with regard to factors associated with cancer incidence
        ii. Maintenance of comparable groups. No large in or out migration during study period; no large drop-outs or differential drop-outs. No differential changes in factors associated with cancer incidence.
        iii. Adequate identification of potential confounders and control of potential confounding by exclusion, stratification, statistical adjustment, other
    ii. Probability of measurement bias *(rate as high/moderate/low)*
        i. Measures of exposure to screening, potential confounders (especially factors related to cancer incidence), and cancer incidence are equally applied between comparison groups
        ii. Measures of exposure to screening, potential confounders, and cancer incidence are valid, including blinding where appropriate.
        iii. Measures of exposure to screening, potential confounders, and cancer incidence are reliable

Follow-up of Randomized Controlled Trial (adapted from the USPSTF Procedure Manual[8])

A. Risk of Bias *(rate overall as high/moderate/low)*
    i. Probability of selection bias *(rate as high/moderate/low)*
        i. Unbiased creation of comparable groups, including adequate randomization, allocation concealment, and equal distribution of potential confounders among both groups
        ii. Maintenance of comparable groups. No large drop-outs or differential drop-outs. Appropriate adherence and minimal contamination or cross-overs.
    ii. Probability of measurement bias *(rate as high/moderate/low)*
        i. Measures of exposure to screening, potential confounders, and cancer incidence are equal between groups
        ii. Measures of exposure to screening, potential confounders, and cancer incidence are valid, including blinding where appropriate
        iii. Measures of exposure to screening, potential confounders, and cancer incidence are reliable
    iii. Potential for confounding *(rate as high/moderate/low)*
        i. Equal distribution of potential confounders among two groups, without changes in group composition throughout follow-up.

Pathologic and Imaging Studies

A. Risk of Bias *(rate overall as high/moderate/low)*
    i. Probability of selection bias and confounding*(rate as high/moderate/low)*

i. No large drop-outs or inadequate follow-up of selected members of study population
ii. If control group present: unbiased creation and maintenance of comparable groups
iii. If control group present: adequate identification of potential confounders and control of potential confounding by exclusion, stratification, statistical adjustment, other

ii. Probability of measurement bias *(rate as high/moderate/low)*
    i. Measures of pathologic or behavioral characteristics are valid, including blinding where appropriate and avoiding differential follow-up
    ii. Measures of pathologic or behavioral characteristics are reliable

Modeling Studies

A. Risk of Bias *(rate overall as high/moderate/low)*
    i. Extent to which assumptions made in the model are transparent and clearly stated *(rate as good/fair/poor)*
    ii. Extent to which assumptions made in the model are backed up with evidence *(rate as good/fair/poor)*
        i. ideally systematically-reviewed evidence that was critical appraised with quality ratings
    iii. Probability for biases in the data used in the model *(rate as good/fair/poor/cannot determine)*
        i. Measurement of outcomes in data used in model are valid and reliable
        ii. Adequate measurement of and control for potential confounders in data used in model
            1. This information should be presented and discussed by authors so that readers can appraise the study.
    iv. Extent to which sensitivity analyses are performed for any uncertain variables *(rate as good/fair/poor)*
        i. ideally probabilistic multivariate sensitivity analyses
    v. Validation: model has been validated using population data different from the population data used to calibrate the model

Supplemental Table 4: Criteria for Evaluating Strength of Evidence

A. Risk of Bias *(rate as high/moderate/low)* (specific criteria listed in Supplemental Table 3)

B. Analysis *(rate as good/fair/poor)* (Ecologic and Cohort, RCT follow-up studies only)
   i. Extent to which the analysis appropriately quantifies overdiagnosis, without inclusion of age groups or time frames that lack the potential to be overdiagnosed, and with appropriate consideration for lead time (i.e., without statistical adjustment for lead time given that these values are derived from models which include overdiagnosed cancers in the estimates of lead time)
   ii. Extent to which the time frame is sufficient to account for the effects of lead time

C. Directness *(rate as good/fair/poor)*
   i. Extent to which the evidence links the screening test directly to health outcomes with minimal assumptions regarding:
      i. The progression of a screen-detected cancer to a cancer that causes morbidity and mortality
      ii. The association of pathologic or behavioral characteristics of a cancer with cancer progression and cancer-related morbidity and mortality

D. External Validity *(rate as good/fair/poor)*
   i. Extent to which study population is similar to US or Western European population in factors that are associated with cancer incidence
   ii. Extent to which the screening situation (e.g., expertise of the screening radiographers, quality of screening facilities, threshold for labeling a result as abnormal) in the study is comparable to the screening situation in the US or Western European population
   iii. Extent to which medical care and risks for competing mortality in the study are similar to medical care in the US or Western European population

E. Precision *(rate as good/fair/poor/cannot determine)*
   i. Confidence interval on magnitude of overdiagnosis should be provided. Width of confidence interval should be narrow.

F. Consistency *(rate as good/fair/poor)*
   i. Degree to which the overdiagnosis measurement from the included studies has a similar magnitude, within the same cancer type and study design

List of Included Studies

Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *New Engl J Med* 2012;367(21):1998-2005.

Ciatto S, Gervasi G, Bonardi R, Frullini P, Zendron P, Lombardi C, et al. Determining overdiagnosis by screening with DRE/TRUS or PSA (Florence pilot studies, 1991-1994). *Eur J Cancer* 2005;41(3):411-415.

Coldman A, Phillips N. Incidence of breast cancer and estimates of overdiagnosis after the initiation of a population-based mammography screening program. *CMAJ* 2013;185(10):E492-8.

Davidov O, Zelen M. Overdiagnosis in early detection programs. *Biostatistics* 2004;5(4):603-613.

De Gelder R, Fracheboud J, Heijnsdijk EAM, den Heeten G, Verbeek ALM, Broeders MJM, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med* 2011;53(3):134-140.

De Gelder R, Heijnsdijk EAM, Van Ravesteyn NT, Fracheboud J, Draisma G, De Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;33(1):111-121.

Dominioni L, Rotolo N, Mantovani W, Poli A, PIsani S, Conti V, et al. A population-based cohort study of chest x-ray screening in smokers: lung cancer detection findings and follow-up. *BMC Cancer* 2012;12.

Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulat R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009;101(6):374-383.

Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Björneld L, et al. Estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005;7(6):258-265.

Duffy SW, Field JK, Allgood PC, Seigneurin A. Translation of research results to simple estimates of the likely effect of a lung cancer screening programme in the United Kingdom. *Br J Cancer* 2014;110(7):1834-1840.

Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen THH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the breast screening programme in England. *J Med Screen* 2010;17(1):25-30.

Falk RS, Hofvind S, Skaane P, Haldorsen T. Overdiagnosis among women attending a population-based mammography screening program. *Int J Cancer* 2013.

Graif T, Loeb S, Roehl KA, Gashti SN, Griffin C, Yu X, et al. Under diagnosis and over diagnosis of prostate cancer. *J Urol* 2007;178(1):88-92.

Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: model estimates of potential benefits and harms. *Ann Intern Med* 2013;158(3):145-153.

Gulati R, Inoue LY, Gore JL, Katcher J, Etzioni R. Individualized estimates of overdiagnosis in screen-detected prostate cancer. *J Natl Cancer Inst* 2014;106(2):djt367. doi: 10.1093/jnci/djt367 [doi]

Gunsoy NB, Garcia-Closas M, Moss SM. Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40 to 49 in the United Kingdom. *Breast Cancer Res* 2012;14(6).

Hazelton WD, Goodman G, Rom WN, Tockman M, Thornquist M, Moolgavkar S, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Math Biosci* 2012;240(1):20-34.

Heijnsdijk EAM, Der Kinderen A, Wever EM, Draisma G, Roobol MJ, De Koning HJ. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *Br J Cancer* 2009;101(11):1833-1838.

Hellquist BN, Duffy SW, Nystrom L, Jonsson H. Overdiagnosis in the population-based service screening programme with mammography for women aged 40 to 49 years in Sweden. *J Med Screen* 2012;19(1):14-19

Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339.

Jorgensen KJ, Zahl PH, Gotzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;9.

Junod B, Zahl PH, Kaplan RM, Olsen J, Greenland S. An investigation of the apparent breast cancer epidemic in France: screening and incidence trends in birth cohorts. *BMC Cancer* 2011;11.

Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian Screening Program. *Ann Intern Med* 2012;157(3):221-222.

Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Tazelaar HD, et al. Five-year lung cancer screening experience: CT appearance, growth rate, location, and histologic features of 61 lung cancers. *Radiology* 2007;242(2):555-562.

Luo D, Cambon AC, Wu D. Evaluating the long-term effect of FOBT in colorectal cancer screening. *Cancer Epidemiol* 2012;36(1):e54; e60.

Martinez-Alonso M, Vilaprinyo E, Marcos-Gragera R, Rue M. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* 2010;12(4).

McGregor M, Hanley JA, Boivin JF, McLean RG. Screening for prostate cancer: estimating the magnitude of overdetection. *Can Med Assoc J* 1998;159(11):1368-1372.

Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* 2014;348:g366. doi: 10.1136/bmj.g366 [doi].

Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control* 2010;21(2):275-282.

Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ* 2013;346.

Olsen AH, Jensen A, Njor SH, Villadsen E, Schwartz W, Vejborg I, et al. Breast cancer incidence after the start of mammography screening in Denmark. *Br J Cancer* 2003;88(3):362-365.

Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res* 2006;8(6).

Pashayan N, Duffy SW, Pharoah P, Greenberg D, Donovan J, Martin RM, et al. Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. *Br J Cancer* 2009;100(7):1198-1204.

Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemägi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014;174(2):269-274.

Peeters PHM, Verbeek ALM, Straatman H, Holland R, Hendriks JHCL, Mravunac M, et al. Evaluation of overdiagnosis of breast cancer in screening with mammography: results of the Nijmegen programmme. *Int J Epidemiol* 1989;18(2):295-299.

Pelzer AE, Colleselli D, Bektic J, Schaefer G, Ongarello S, Schwentner C, et al. Clinical and pathological features of screen vs non-screen-detected prostate cancers: is there a difference? *BJU Int* 2008;102(1):24-27.

Pinsky PF. An early- and late-stage convolution model for disease natural history. *Biometrics* 2004;60(1):191-198.

Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E. Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res* 2012;14(1).

Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;45(18):3166-3171.

Seigneurin A, Francois O, Labarere J, Oudeville P, Monlong J, Colonna M. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ (Online)* 2012;344(7839).

Sobue T, Suzuki T, Matsuda M, Kuroishi T, Ikeda S, Naruke T. Survival for clinical stage I lung cancer not surgically treated: comparison between screen-detected and symptom-detected cases. *Cancer* 1992;69(3):685-692.

Sone S, Nakayama T, Honda T, Tsushima K, Li F, Haniuda M, et al. Long-term follow-up study of a population-based 1996-1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography. *Lung Cancer* 2007;58(3):329-341.

Svendsen AL, Olsen AH, Von Euler-Chelpin M, Lynge E. Breast cancer incidence after the introduction of mammography screening: what should be expected? *Cancer* 2006;106(9):1883-1890.

Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics* 2008;64(1):10-19.

Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Stat Med* 2006;25(16):2846-2866.

Veronesi G, Maisonneuve P, Bellomi M, Rampinelli C, Durli I, Bertolotti R, et al. Estimating overdiagnosis in low-dose computed tomography screening for lung cancer: a cohort study. *Ann Intern Med* 2012;157(11):776-784.

Wu GHM, Auvinen A, Maattanen L, Tammela TLJ, Stenman UH, Hakama M, et al. Number of screens for overdetection as an indicator of absolute risk of overdiagnosis in prostate cancer screening. *Int J Cancer* 2012;131(6):1367-1375.

Yankelevitz DF, Kostis WJ, Henschke CI, Heelan RT, Libby DM, Pasmantier MW, et al. Overdiagnosis in chest radiographic screening for lung carcinoma: frequency. *Cancer* 2003;97(5):1271-1275.

Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 2006;332(7543):689-691.

Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. *Tidsskr Nor Legeforen* 2012;132:414-417.

Zahl PH, Strand BH, Maeshlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;328(7445):921-924.

Zappa M. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence screening pilot study. *Ann Oncol* 1998;9(12):1297-1300.