Appendix

In 2016, the U.S. Preventive Services Task Force (USPSTF) published its current guidelines for breast cancer screening. The updated recommendations were based on an analysis of the data from seven randomized controlled trials of breast cancer screening in asymptomatic women age 39 and older. An eighth trial (i.e., the Edinburgh trial conducted in the United Kingdom) was excluded from the USPSTF evidence analysis due to significant concerns over inadequate randomization. Background details regarding each of the included trials are presented in **Table 1**. All of the included studies were rated as having "fair" quality of evidence.[1]

Table 1. Randomized Clinical Trials of Breast Cancer Screening Used For Evaluation of Benefits and Harms of Breast Cancer Screening.

Trial	Year trial began	Number of women	Ages at enrollment	Years of longest follow up	Country	New data since 2009 USPSTF guidelines
New York Trial (HIP Trial)	1963	60,495	40-64	18	United States	No
Malmö Trial (MMST I and MMST II)	1976	42,283	43-69	15	Sweden	No
Swedish Two-County Trial	1977	133,065	40-70	20	Sweden	Yes
Canadian Trials* (CNBSS-1 and CNBSS-2)	1980	89,835	40-49 (CNBSS-1) 50-59 (CNBSS-2)	25	Canada	Yes
Stockholm Trial	1981	60,117	40-64	11	Sweden	No
Göteborg Trial (Gothenburg Trial)	1982	49,924	40-59	12	Sweden	No
Age Trial	1991	160,921	39 - 41	17	United Kingdom	Yes

*Two Canadian trials (CNBSS-1 and CNBSS-2) are combined in Table 1 above.

[1] RCTs included in the USPSTF analysis could be rated as "Good," "Fair," or "Poor" quality. Definitions for each are provided below.

Good: Comparable groups are assembled initially and maintained throughout the study (followup at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.

Fair: Studies are graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred in followup; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.

Poor: Studies are graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat is lacking.

The results of the USPSTF analysis are presented in **Table 2.** As reported in the USPSTF meta-analysis, the relative risk (RR) of breast cancer mortality for women in each of the following age groups, with screening was:

- **39 to 49 years**: RR of death = 0.92 (95% CI, 0.75 to 1.02; Not significant; based on 9 trials) Equivalent to 3 deaths prevented per 10,000 women over 10 years
- **50 to 59 years**: RR of death = 0.86 (95% CI, 0.68 to 0.97; Significant; based on 7 trials) Equivalent to 8 deaths prevented per 10,000 women over 10 years
- **60 to 69 years**: RR of death = 0.67 (95% CI, 0.54 to 0.83; Significant; based on 5 trials) Equivalent to 21 deaths prevented per 10,000 women over 10 years
- **70 to 74 years**: RR of death = 0.80 (95% CI, 0.51 to 1.28; Not significant; based on 3 trials Equivalent to 13 deaths prevented per 10 000 women over 10 years

All-cause mortality (i.e., all deaths among study participants, regardless of the cause) was not reduced with screening irrespective of whether trials were analyzed in combined or separate age groups.

A meta-analysis of the incidence of advanced breast cancer outcomes reported in the screening trials demonstrated that advanced stage diagnoses (i.e., Stage III and IV[2]) was reduced for women aged 50 years or older (RR, 0.62 [CI, 0.46 to 0.83]) (3 trials) but not those aged 39 to 49 years (RR, 0.98 [CI, 0.74 to 1.37]) (4 trials). However, these findings were based on fewer trials eligible for the analysis than the mortality estimates, and they differ from studies of population trends that show little to no reductions in advanced breast cancer after the introduction of screening mammography.

Table 2. Analyses of the Mammography Screening Randomized Control Trials.

Age at screening	Relative risk of death	95% Confidence Interval	Number of trials	Number of deaths avoided per 10,000 women screened over 10 years (95% CI)	All causes of mortality	Relative risk of advanced breast cancer (95% CI)
39-49	0.92	0.73 to 1.02	9	3 (0 to 9)	No change	0.98 (0.74 to 1.37) 4 trials
50-59	0.86	0.54 to 0.83	7	8 (2 to 17)	No change	ND
60-69	0.67	0.55 to 0.91	5	21 (11 to 32)	No change	ND
70-74	0.80	0.51 to 1.28	3	13	No change	ND
>50 (combining all data for women older than 50)	ND	ND	ND	ND	No change	0.62 (0.46 to 0.83) 3 trials

Note: A relative risk of 1 means that screening did not improve survival or reduce the risk of developing advanced breast cancer, which is defined as stage III or IV cancer or tumors greater than 40 mm with positive lymph nodes. A relative risk less than 1 means that women who received mammography screening had a reduced risk. Relative risk values close to 1 mean that there is not a significant difference in risk between women who were randomly assigned to be screened and those were not. ND=not determined.

[2] Defined as regional or metastatic, size 50 mm or greater, or having four or more positive lymph nodes (Stage III or IV by the AJCC TNM system).

A similar analysis of the above-described randomized controlled trials of screening mammography was conducted by the Cochrane Collaboration and published most recently in 2013. Similar to the USPSTF analysis, all eligible trials were assessed for key characteristics of the trial, study quality and risk for bias.[3]

Of the included trials, three were classified as adequately randomized (Canada, Malmö and UK age trial) and the other four as sub-optimally randomized (Göteborg, New York, Stockholm, Two County), as was also the extension of the Malmö trial, MMST II.

Similar to the USPSTF analysis, the Edinburgh trial was not adequately randomized and cannot provide reliable data and was therefore not included in the meta-analysis.

The Cochrane investigators examined the pooled estimates for the trials with adequate randomization and those with suboptimal randomization together, as well as in separate subgroup analyses by quality of randomization. Estimates for outcome data were examined at 7 years and 13 years follow-up. Relative risk ratios for specific analyses for data at 13 years for breast cancer mortality are presented in Table 3 below.

Among the three trials with adequate randomization (i.e., medium-quality randomization) no statistically significant reduction in breast cancer mortality at 13 years was observed (relative risk (RR) 0.90, 95% confidence interval (CI) 0.79 to 1.02). The four trials with suboptimal randomization showed a statistically significant reduction in breast cancer mortality with an RR of 0.75 (95% CI 0.67 to 0.83). The RR for all seven trials combined was 0.81 (95% CI 0.74 to 0.87).

Table 3: Breast Cancer Mortality – 13 Years Follow-up

Pooled Analyses of RCT's of Screened vs. Unscreened	Relative Risk (Fixed) 95% CI
All ages	
• Breast cancer deaths, 13 years follow up (pooled across all trials) (N=7)	0.81 (0.74 to 0.87)*
• Medium-quality trials (N=3)	0.90 (0.79 to 1.02)
• Low-quality trials (N=4)	0.75 (0.67 to 0.83)*
Age < 50	
• Breast cancer deaths, 13 years follow up (pooled across all trials) (N=6)	0.89 (0.72 to 1.10)
• Medium-quality trials (N=2)	1.03 (0.77 to 1.38)
• Low-quality trials (N=4)	0.77 (0.67 to 1.04)
Age > 50	
• Breast cancer deaths, 13 years follow up (pooled across all trials) (N=5)	0.76 (0.66 to 0.86)*
• Medium-quality trials (N=2)	0.94 (0.77 to 1.15)
• Low-quality trials (N=3)	0.64 (0.54 to 0.78)*

*Reflects statistically significant reduction.

[3] Key elements of included studies that were examine in determining the risk of bias included: population studied, comparability of groups, assignment of cause of death, randomization and blinding procedures and exclusions after randomization, and the likelihood of selection bias.