MODEL DESCRIPTION AND CHOICE OF MODELS

ITEM	QUESTION	APPRAISAL
Model description	Are the models used in the collaborative modeling effort identified clearly?	ADEQUATE: There is a clear and complete description of the models used in an accepted and consistent terminology. INADEQUATE: It is not possible to figure out what models were used; or contradictory statements are made what models were used.
Concordance of models and implications	Are the important differences between the models described and the implications of these differences discussed?	ADEQUATE: Important differences between models are described and implications discussed; or evidence is presented that there are no important differences. INADEQUATE: There is no discussion of whether there are differences among models; or there are differences identified but the implications of the differences are not discussed.
Time horizon considering the screening question to be addressed	Is the time horizon for the collaborative modelling appropriate for the screening question addressed?	ADEQUATE: The time horizon for the collaborative modeling of the model is appropriate for the screening question. INADEQUATE: The time horizon is not stated; or the time horizon is not appropriate for the screening question addressed.
Calibration	Were the models used in the collaborative modeling effort calibrated (tested) against independent data? Were the data used to calibrate the models relevant for the US prevention population?	ADEQUATE: All of the models were calibrated against SEER (preferred); or models were calibrated using clinical trial data. POSSIBLY ADEQUATE: The models used in the collaborative effort were reported to be calibrated against a representative non-US population that has applicability to the US population. INADEQUATE: Calibration of the models against independent data is not mentioned; or calibration of the models was not done.

BENEFIT OF SCREENING

ITEM	QUESTION	APPRAISAL
How benefit of screening arises	Does the report describe how the benefit of screening is conceptualized (e.g., stage shift and curative treatment or mortality reduction due to decreased relative risk) for each model?	ADEQUATE: The way that the benefit of screening is conceptualized is described for each model. INADEQUATE: The way that the benefit of screening is conceptualized is not described for one or more models.

DATA

ITEM	QUESTION	APPRAISAL
Data identification/Benefits of the preventive service	Are the sources of data that determine the benefit of screening intervention identified? Has the process of selecting the data on key parameters related to the benefit of the preventive intervention been systematic?	ADEQUATE: The sources of data on the benefits of the preventive intervention are identified and the source is a good quality systematic review (preferred); or the sources of data on benefits are not a systematic review but the primary sources are justified and credible. INADEQUATE: The sources of data on the benefit of the preventive intervention are not identified; or data are drawn from a poor systematic review; or the data sources are not justified and credible.
Harms considered	Is a rationale for the selection of harms to be considered provided? Are there important harms that were not considered?	ADEQUATE: The most important harms are included and a rationale is provided. POSSIBLY ADEQUATE: Some but not all of the most important harms are considered. INADEQUATE: No rationale is provided for the selection of harms to be considered; or important harms are not considered.

Data identification/harms	Are the sources of data on the harms of the preventive intervention identified? Has the process of selecting the data on key parameters related to the harms of preventive intervention been systematic?	ADEQUATE: The sources of data on the harms of the preventive intervention are identified and the source is a good quality systematic review (preferred); or the sources of data on the harms are not drawn from a systematic review but the primary sources are justified and credible. INADEQUATE: The sources of data on the harms of the preventive intervention are not identified; or the data are drawn from poor systematic review; or the data sources about harms are not justified and credible.
Assessment of parameter uncertainty	Has deterministic sensitivity analysis been done? Is the choice of parameters for the deterministic sensitivity analysis explained? Is the explanation credible?	ADEQUATE: Deterministic sensitivity analysis is done for all of the parameters that might affect the model results as they relate to the benefits and harms of screening; or only some parameters are used in the deterministic sensitivity analysis but the reasons for their choice are given and they make sense. Results of the deterministic sensitivity analysis are reported. POSSIBLY ADEQUATE: Only some parameters are subjected to deterministic sensitivity analysis but the reasons for the choice of these parameters are not given or do not make sense. INADEQUATE: Deterministic sensitivity analysis is not done; or deterministic sensitivity analysis is described but the results are not reported.
Assessment of parameter uncertainty	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly? Are the ranges justified?	ADEQUATE: The ranges used for sensitivity analysis are stated clearly and the reasons for the ranges are explained and make sense. INADEQUATE: The ranges used for sensitivity are not described; or the ranges used for sensitivity analysis are not explained; or the ranges used for the sensitivity analysis don't make sense.

Assessment of parameter uncertainty	Has probabilistic sensitivity analysis been done? If not, has this been justified?	ADEQUATE: Probabilistic sensitivity analysis is done (preferred); or probabilistic sensitivity analysis is not done but the reasons are given and make sense. INADEQUATE: Probabilistic sensitivity analysis is not considered.
Heterogeneity [possible differences in model derived benefits and harms for subgroups]	Has heterogeneity for the screening question been considered?	ADEQUATE: Heterogeneity is considered and important subgroups are included. POSSIBLY ADEQUATE: Heterogeneity is considered but at least some important subgroups are not included; or heterogeneity is not considered but reasons for doing this are given and make sense. INADEQUATE: Heterogeneity is likely present but is not considered.

VALIDATION AND CONSISTENCY

ITEM	QUESTION	APPRAISAL
Cross-validation	Is a comparison of the results of the different models used in the collaborative modeling effort presented? Are the results consistent with each other? If not, are the reasons for the inconsistency explained? Do differences in model inputs and assumptions explain the differences in outcomes?	ADEQUATE: The results of the different models used in the collaborative effort are compared as they pertain to the main issues addressed and the results are generally consistent; or the results are compared and are inconsistent but the reasons for inconsistency are explained and are credible. POSSIBLY ADEQUATE: The results of the different models are compared and are inconsistent and an attempt is made to reconcile differences but the attempt is not wholly convincing. INADEQUATE: The results of the different models are not compared. The results are not consistent and no attempt is made to explain the reasons for inconsistency in the results.
Conclusions	Are overall conclusions robust across the models?	ADEQUATE: The overall conclusions are robust across models (preferred); or the overall conclusions are not robust across models but the reasons are explained and are credible. INADEQUATE: Overall conclusions are not robust across models and the reasons for the difference are not explained; or it is not possible to determine whether the overall conclusions are robust across models.

External consistency	Was a comparison of the ability of the models to reproduce external empiric data done?	ADEQUATE: The ability of all the models to reproduce external data is evaluated. INADEQUATE: The ability of the models to reproduce external data is not evaluated if such data is available and relevant;
External consistency	If yes, was the comparison of the ability of the models to reproduce external empiric data presented? Was the fit of the models to empiric data good? If the fit of any of the models to empiric data was not good, have the reasons been explained? If the fit of any of the models to empiric data was not good, was a sensitivity analysis of the overall results done excluding the model(s) with poor fit?	ADEQUATE: The fit of all of the models to the data is good (preferred); or the ability to reproduce external data is not good for one or more included models but the explanation given for the lack of fit to external data is credible or the modeler make a convincing case that the lack of fit does not affect conclusions with regard to the preventive service. INADEQUATE: The ability to reproduce external data is not good for one or more included models and the explanation given for the lack of fit to external data is not credible or no explanation is provided.

COMMUNICATION

ITEM	QUESTION	APPRAISAL
Transparency	Is a detailed description of all of the models used in the collaborative effort available publicly?	ADEQUATE: A detailed description of all the models is available publicly. INADEQUATE: There is no publicly available description of one or more of the models; or the description of one or more of the models is superficial and/or incomplete.
Interpretability	Are the results of the collaborative modeling effort clear and easily interpretable?	ADEQUATE: The results of the collaborative modeling effort are clear and easily interpretable. INADEQUATE: The results of the collaborative modeling effort are difficult to understand and interpret.

CONFLICT OF INTEREST

ITEM	QUESTION	APPRAISAL
Financial conflict or non-financial bias	Is the source of funding for collaborative model effort identified? Was the source of funding a for-profit entity? If yes, were steps taken to assure that the collaborative modeling results were free of influence of the funding entity? Is there reason to suspect significant non-financial bias?	ADEQUATE: Source of funding is identified and there are no significant financial conflicts identified; or a mitigation plan supplied for any significant financial conflicts. There is no obvious non-financial bias. INADEQUATE: Source of funding is not identified; or significant conflicts revealed without any mitigation plan; or there is reason to suspect significant non-financial bias which has not been mitigated.