Differences between CRC-AIM and CRC-SPIN

While CRC-AIM was inspired by CRC-SPIN, there are several differences. Almost all the estimated parameters between CRC-AIM and CRC-SPIN differ as the two models use different approaches for model calibration. Namely, while CRC-SPIN used a Bayesian calibration method, we utilize a machine learning-based empirical calibration approach as explained in more detail below. The differences in estimated parameters led to different predicted outcomes by the models (e.g., sojourn time, dwell time, adenoma size distribution, life-years gained, incidence reduction, and mortality reduction), further highlighting the importance of including structural or parametric sensitivity analysis within natural history modeling to ensure robust conclusions.

In addition to the differences in parameter values, there are three major structural differences between the two models. First, CRC-AIM uses yearly cycles for adenoma generation and transition into preclinical cancer, as opposed to a continuous-time model employed by CRC-SPIN. While the use of yearly cycles precludes the generation of multiple adenomas or transition from adenoma to clinical CRC within the same year, such events are rare as quantified by the transition probabilities reported before.[25] Second, CRC-AIM models survival from CRC using cause-specific survival, rather than the relative survival estimates from an analysis of SEER 1975-2003 data.[8] A presentation by NCI notes the limitations associated with using relative survival and suggests the use of cause-specific survival for accurate representation of post-diagnosis survival.[26] Additionally, CRC-AIM considers the improvement in survival since 2003 through advancements in cancer treatments by applying a 7% hazard reduction, which is estimated using the 5-year survival between periods 2000-2003 and 2010-2019 from SEER. No survival improvement was modeled in CRC-SPIN. Third, the CRC size and stage estimation differs between the two models. The stage at clinical detection in CRC-AIM is modeled as a function of anatomic subsite location (proximal colon, distal colon, and rectum), age-group at diagnosis, and sex. CRC-SPIN only included age and sex in modeling this variable. CRC-AIM models the size at clinical detection based on location and stage at clinical detection using SEER 2010-2015 data, limited to cases diagnosed at ages 20-50, where the rationale has been provided in the main manuscript. On the other hand, CRC-SPIN used 1975-1979 SEER data for this purpose. CRC-AIM models stage-shift due to screening by randomly assigning a stage at screening detection based on the distribution of stage at clinical detection but imposing the restriction that the newly sampled stage must not be worse than the stage at clinical detection. The modeling of stage shift in CRC-SPIN is not reported in detail.

While CRC-AIM is inspired by CRC-SPIN, it is worth noting that CRC-SPIN has evolved, and CRC-AIM closely resembles the updated version of CRC-SPIN, except for the transition from adenoma to preclinical cancer of CRC-AIM is adapted from the original CRC-SPIN.

Finally, the approach to calibration substantially differs between CRC-AIM and CRC-SPIN. CRC-AIM utilizes a machine learning based framework as an emulator to estimate the targets for an input vector. On the other hand, CRC-SPIN employs an incremental mixture approximate Bayesian computation (IMABC) approach, which is a sequential, rejection-based method that iteratively adds new points near regions that are close to the calibration targets then decreases the tolerance intervals. The algorithm stops when a sufficient number of accepted points is obtained. Posterior estimates are then calculated by weighting the final set of accepted points. Each iteration of the IMABC approach requires simulation of calibration targets, thus this approach can be computationally intensive. However, the IMABC approach allows the use of a large number of

calibration targets (CRC-SPIN calibrated to a total of 40 targets) whereas CRC-AIM's emulator was limited to 8 calibration targets. The performance of our machine learning based framework using a larger number of calibration targets requires further research. Additionally, CRC-AIM classifies the calibration targets using a hierarchical order, whereas CRC-SPIN fits to all targets in each iteration.