PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A novel microsimulation model of tobacco use behaviors and outcomes: calibration and validation in a US population
AUTHORS	Reddy, Krishna; Bulteel, Alexander; Levy, Douglas; Torola, Pamela; Hyle, Emily; Hou, Taige; Osher, Benjamin; Yu, Liyang; Shebl, Fatma; Paltiel, A David; Freedberg, Kenneth; Weinstein, Milton; Rigotti, Nancy; Walensky, Rochelle

VERSION 1 – REVIEW

REVIEWER	Tam, Jamie Yale University School of Public Health I have an appointment in the Department of Health Policy and Management, Yale School of Public Health, where one of the co-authors, A David Paltiel, also has an appointment.
REVIEW RETURNED	03-Sep-2019

GENERAL COMMENTS	This study advances tobacco simulation modeling by incorporating relapse behaviors--an improvement over previous models. The paper is well-written and provides relevant methodological details with data presented clearly. In general, this is a novel and worthwhile contribution to the literature, and indeed having multiple validated models of smoking behaviors could help inform future policy. Several changes would improve the paper from its current form: Introduction: - The third paragraph describes the shortcomings of previous tobacco models, but does not make a good argument as to why we need a model that incorporates smoking relapse. Why is a relapse model more useful? What are the additional applications that could not be addressed by previous models? The last sentence states that this model could "inform clinical and public health policy", but it's not clear how. While some of this is described in the Discussion section, the Introduction on its own does not make a compelling argument for this as a research need, and this sells the study short. Methods: - Recommend adding a new paragraph with the subheader "Smoking definitions," that describes how "never smoker", "current smoker", "former smoker", and "recent quitter" are each defined. The paragraph should furthermore note whether the definitions differ from other models or data sources, and why the authors defined them as such.

	- Page 8, lines 12-17: "Those who quit smoking retain the all-cause mortality…until they have remained abstinent for a defined period of time (e.g., five years)…" Why five years? The CISNET models apply former smoker mortalities based on a 2-year abstinence period. - Page 8, lines 23-26: "Upon an individual's death, the next simulated person enters the model." Why did the authors decide to have a constant population size in their model? Why not incorporate historical birth rates, historical population sizes? - Table 2: "Clinical events" does not seem like the appropriate header for what are specifically "Mortality events". Suggest rephrasing. - Explain why older data were used for this study (1997-2009 NHIS data with mortality follow-up through 2011) rather than newer data that could reflect more recent changes in smoking behaviors (NHIS mortality follow-up data are available through 2015). - Page 11, footnote e: "CISNET-derived former smoker mortality rates are often lower than CISNET-derived never smoker mortality rates for the 1950 birth cohort- relationship with questionable face validity." Has this been confirmed with the authors by CISNET? If so, consider citing the correspondence. Results: - Page 15, lines 43-46: "These multipliers fall within the published range of values…" This is commentary that belongs in the Discussion section, not the Results section. - Page 16, lines 21-24: "These are similar to the median life expectancies for…" Please move commentary to the Discussion section. Discussion: - Page 19, line 44:-50: These sentences were confusing. Why would STOP formally label current 'some-day' smokers as former smokers? These are individuals who are reporting as current smokers in the NHIS. If someone responds to the survey question that they currently smoke "some days", then they certainly have not been abstinent long enough to self-identify as former smokers-- and they are not considered former smokers in the NHIS. Please explain in the Methods section. Figure 1: Suggest changing "smoker who recently stopped" to "Recent quitter" and using this term (or another brief term) consistently throughout the paper. Supplementary Table 1: Why was 5 years chosen for the permanent quit transition time? Did the authors consider other durations? It would be helpful to see some type of sensitivity analysis around this assumption. Supplement - Mortality Inputs: The NHIS Linked mortality follow-up only collects smoking behavior information at baseline, and not at follow-up. This means that someone categorized as a current smoker at baseline, may well have been a former smoker by time of death. Did the authors perform any type of right censoring to avoid the potential for misclassification?

REVIEWER	Chris Kypridemos & Vincy Huang University of Liverpool

REVIEW RETURNED	11-Nov-2019

GENERAL COMMENTS	Thank you for giving me the opportunity to review this interesting, and methodologically sound paper. It describes the calibration and validation of a Monte Carlo microsimulation model of tobacco smoking. The model outputs are estimates of smoking prevalence and all-cause mortality among US population and the microsimulation can presumably be used in estimating the effectiveness of tobacco control interventions. In the model, the authors incorporated smoking relapse to project smoking status although this is not unique or as previous microsimulations also include relapse probabilities in their calculations. What it is certainly rarer (and possibly unique??) in discrete time microsimulation is that they do so on a monthly rather than annual basis (although they use monthly probabilities converted from annual probabilities, so seasonality is not captured). I am happy to recommend this manuscript for publication after the authors adequately address my comments below. Major: Figure 3 and external mortality validation: It is important to additionally validate against ex-smoker mortality (or overall mortality). The inclusion of relapse times and the calibrated cessation probabilities have biggest impact on ex-smokers' mortality and this crucial sub-population needs to be included in the external mortality validation preferably directly or indirectly by using overall mortality not stratified by status (that includes never, ever, and current smokers). P18 last paragraph (cont. in p19) it completely ignores calibration drift, a well-documented phenomenon. My understanding is that the current implementation of STOP does nothing to prevent calibration drift, therefore, the model is unsuitable "to predict future tobacco use" as the authors claim. Please discuss the fact that calibration of cessation rates may (over)compensates for other inaccuracies in model inputs or model structure. Your mentioned sensitivity analysis where you simultaneously calibrated initiation rates and the alignment of your multiplier with empirical evidence partially addresses the issue. Calibration on historic data or other models is not panacea, and while the potential future uses of the model are enthusiastically

	discussed, the limitations and implicit assumptions they involve are understated Minor: The introduction is not completely relevant to the main body of the paper which is methodological in nature. There are themes like e-cigs utilisation or that “the decline has not been seen in all segments of society” that are only remotely relevant if at all to the presented research. Please provide a brief description of the CISNET model which you use to calibrate STOP (or move the one from the supplement). It is not entirely clear to me why you chose the term internal validation instead of cross validation for your comparison with the CISNET model. Do they share common inputs? I find table 2 very confusing to the extend that I cannot offer a suggestion for improvement. Splitting the table into 3 tables (one for internal and 2 for external validation) may be a good starting point. P12 L6. Other microsimulations in the past have included relapse probability in their calculations. Please tune down the phrase “A novel aspect of the STOP model”. METHODS: Authors clearly described definitions of transient quit attempts and sustained abstinence. However, they were only mentioned in the middle part of method section. Authors could define those terminologies in the first few paragraphs to give readers a clear idea. P18 L30-46. The approach that is described here to assess potential efficacy of interventions requires strong, not data-driven assumptions of the modelled intervention on reducing relapse probability in the long run. The way it is described it compromises the usefulness of the model in assessing efficacy. I am happy to review a revised version.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

6. The third paragraph describes the shortcomings of previous tobacco models, but does not make a good argument as to why we need a model that incorporates smoking relapse. Why is a relapse model more useful? What are the additional applications that could not be addressed by previous models? The last sentence states that this model could "inform clinical and public health policy", but it's not clear how. While some of this is described in the Discussion section, the Introduction on its own does not make a compelling argument for this as a research need, and this sells the study short.

Response: We have modified the third paragraph of the Introduction to describe the usefulness of a model that includes relapse, the applications that could not be addressed by previous models, and the implications for clinical and public health policy and trial design.

Introduction, pages 5-6: “A current challenge of projecting longer-term clinical and economic outcomes of short-term tobacco cessation studies lies in capturing the many smoking quit attempts and relapses [Hughes et al., Nicotine Tob Res 2014; Chaiton et al., BMJ Open 2016]. A new model that intentionally examines relapse would extend trial results by projecting outcomes beyond the time horizon of trials, when many relapses occur. Our objective was to develop, calibrate, and validate a novel, individual-level microsimulation model that directly addresses the mechanics of smoking initiation, cessation, and relapse, and the associated clinical outcomes. The intended applications of the model include projecting the downstream impact of clinical and public health policy decisions and informing the design of tobacco treatment trials.”

7. Recommend adding a new paragraph with the subheader "Smoking definitions," that describes how "never smoker", "current smoker", "former smoker", and "recent quitter" are each defined. The paragraph should furthermore note whether the definitions differ from other models or data sources, and why the authors defined them as such.

Response: We have added a new paragraph, as recommended, and modified another paragraph.

Methods, page 9:
“Smoking definitions
Similar to NHIS and the Cancer Intervention and Surveillance Modeling Network (CISNET, which used NHIS data), we defined Never Smokers as those who had smoked <100 cigarettes in their lifetime. Among others (ever smokers), NHIS defined current smokers as those who reported currently smoking every day or some days. NHIS considered ever smokers who reported no smoking at the time of interview to be former smokers, regardless of the duration of abstinence. CISNET considered former smokers to be those who had quit smoking at least two years prior to interview; those with a shorter period of abstinence were still considered current smokers.

To better distinguish relapse and mortality risks among those with short-term or long-term abstinence, the STOP model includes three states for those who have ever smoked: 1) Current Smoker; 2) Recent Quitter (short-term abstinence); 3) Former Smoker (long-term abstinence) (figure 1). This enables a differentiation between: 1) transient quit attempts: transition from the Current Smoker state to the Recent Quitter state, with a relatively high rate of early relapse back to the Current Smoker state; and 2) sustained abstinence: transition from the Recent Quitter state to the Former Smoker state, with a lower rate of later relapse back to the Current Smoker state.”

8. Page 8, lines 12-17: "Those who quit smoking retain the all-cause mortality…until they have remained abstinent for a defined period of time (e.g., five years)…" Why five years? The CISNET models apply former smoker mortalities based on a 2-year abstinence period.

Response: The Reviewer's comment highlights the inconsistency in the literature regarding definition of "former smoker." The CISNET model used a 2-year abstinence period (Holford et al., Am J Prev Med 2014) but did not cite data to support this choice. NHIS, meanwhile, did not include any abstinence period. Since it was impossible to choose an abstinence period that would match both CISNET and NHIS, we elected to apply a definition of 5 years, a value that appeared to be best supported by contemporary data (Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013). We have revised the text to clarify our reasoning:

Supplement, page 4: "This reflects contemporary studies in which former smokers were defined as those who had not smoked in the previous five years and data from large US cohort studies that indicate that the all-cause mortality risk in men who quit smoking does not fall below that of current smokers until five years of abstinence. [Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013]"

We also applied this 5-year abstinence period in our previous studies that used a different model (Reddy et al., J Infect Dis 2016; Reddy et al., JAMA Intern Med 2017). Because the "STOP model structure" section may not be the best place to indicate this user-defined criterion, we have removed it from the sentence cited by the Reviewer.

Methods, STOP model structure, page 10: "Those who quit smoking retain the all-cause mortality probabilities of current smokers until maintaining abstinence for a defined period of time (e.g., five years), after which the mortality probabilities decline."

Given the uncertainty around definitions of "former smoker", we performed a sensitivity analysis around the 5-year abstinence period. We now describe this in the Supplement.

Supplementary Table 1, footnote b: "This five-year abstinence period was based on data showing mortality risks by years since smoking cessation. [Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013] We performed an analysis in which we assumed that mortality risks decreased to "Former Smoker" levels immediately upon quitting smoking, but the model-generated results were not a better fit to NHIS data (results not shown)."

9. Page 8, lines 23-26: "Upon an individual's death, the next simulated person enters the model." Why did the authors decide to have a constant population size in their model? Why not incorporate historical birth rates, historical population sizes?

Response: Because the outcomes of interest in our model are "averages" (prevalence, mortality, life expectancy) rather than absolute numbers, historical birth rates and population sizes are not needed. Instead, given Monte Carlo variability, we apply a constant population size that is large enough to provide stable "average" estimates but not so large as to be excessively burdensome from a computation standpoint. We have clarified this.

Methods, page 10: "We use a constant simulated population size of one million to obtain stable estimates of these "average" outcomes of interest."

10. Table 2: "Clinical events" does not seem like the appropriate header for what are specifically "Mortality events". Suggest rephrasing.

Response: We have made the suggested change.

Table 2: “Mortality” (instead of “Clinical Events”)

11. Explain why older data were used for this study (1997-2009 NHIS data with mortality follow-up through 2011) rather than newer data that could reflect more recent changes in smoking behaviors (NHIS mortality follow-up data are available through 2015).

Response: We used NHIS data through 2009 because those were the data used in the CISNET studies, which were our comparator in cross-validation exercises. At the time we conducted our analyses, linked mortality follow-up data were available through 2011. Regardless, the objective of our analysis was to calibrate our model and demonstrate validity using historical cohorts, rather than informing current epidemiology or projecting future smoking prevalence or disease burden, which we aim to do in future analyses and is the motivation for building this model. We have modified the description in the Methods.

Methods, page 11: “For the initial cross-validation exercise, we used data from CISNET modeling studies, which were derived from NHIS through 2009 and were stratified by birth cohort…”

Methods, page 17: “We used NHIS data through 2009 because those were the data used in the CISNET studies, which were our comparator in cross-validation exercises.”

12. Page 11, footnote e: "CISNET-derived former smoker mortality rates are often lower than CISNET-derived never smoker mortality rates for the 1950 birth cohort- relationship with questionable face validity." Has this been confirmed with the authors by CISNET? If so, consider citing the correspondence.

Response: These data appear on the CISNET website (https://resources.cisnet.cancer.gov/projects/#shg/tcpd/efig_4/opts/sex=3;render_method=table;interval=10;show_all_series=true;current_series=1970;show_reference=false). eFig 4 shows death rate per year. For example, in 1990, the mortality rate for 40-year-old male former smokers is reported as 0.0016 per year, and the mortality rate for 40-year-old male never smokers is 0.0024 per year. In 2000, the mortality rate is reported as 0.0032 per year for 50-year-old male former smokers, compared to 0.0041 per year for 50-year-old male never smokers. We have modified the footnote to clarify.

Table 2, footnote e: “For the 1950 birth cohort, some CISNET-derived former smoker mortality rates are lower than CISNET-derived never smoker mortality rates – a counterintuitive relationship otherwise unexplained. We therefore adapted former smoker mortality multipliers for the cross-validation from Thun et al.”

13. Page 15, lines 43-46: "These multipliers fall within the published range of values…" This is commentary that belongs in the Discussion section, not the Results section.

Response: We have removed this statement from the Results section. A similar statement already appears in the Discussion section (page 22, “The cessation multipliers that provided the best fits to empirical data are in line with published data regarding the number of quit attempts required before sustained abstinence is achieved”).

14. Page 16, lines 21-24: "These are similar to the median life expectancies for…" Please move commentary to the Discussion section.

Response: We have moved the statement from the Results section to the Discussion section.

Discussion, page 24: “On the other hand, STOP model-generated life expectancies were similar to the median life expectancies for 30-year-old smokers reported by Jha et al. (also derived from NHIS data): 77 years for women and 72 years for men.”

15. Page 19, line 44-50: These sentences were confusing. Why would STOP formally label current 'some-day' smokers as former smokers? These are individuals who are reporting as current smokers in the NHIS. If someone responds to the survey question that they currently smoke "some days", then they certainly have not been abstinent long enough to self-identify as former smokers-- and they are not considered former smokers in the NHIS. Please explain in the Methods section.

Response: To provide a more detailed explanation, we have modified the Methods and the Discussion.

Methods, page 10: “For the purpose of model output displays, those in the Recent Quitter state are considered “former smokers.””

Discussion, pages 23-24: “In an external validation exercise, the STOP model projection for never smoker prevalence from 1998 to 2009 was slightly lower than that reported by NHIS, and the STOP model projection for former smoker prevalence was slightly higher than NHIS data. In NHIS, former smokers were self-defined but on average had been abstinent for over a decade. NHIS considered those who smoked “some days” to be current smokers, though some of them may have been in the midst of a short-duration quit attempt. STOP model output formally labels these people, who may be in the Recent Quitter state, former smokers but assigns them the mortality risks of current smokers (until a defined period of abstinence). STOP reflects monthly quitting and relapsing behaviors whereas NHIS is an annual cross-sectional survey. Thus, one would expect the STOP model to report a higher prevalence of former smokers than NHIS, as seen in our results.”

16. Figure 1: Suggest changing "smoker who recently stopped" to "Recent quitter" and using this term (or another brief term) consistently throughout the paper.

Response: We have changed the term to “Recent Quitter” in Figure 1 and throughout the manuscript and supplement.

17. Supplementary Table 1: Why was 5 years chosen for the permanent quit transition time? Did the authors consider other durations? It would be helpful to see some type of sensitivity analysis around this assumption.

Response: Please see our response to Comment #8, above. In addition, we have described a sensitivity analysis we performed. Changing the time until mortality risk decreases to that of a Former Smoker is a proxy for the permanent quit transition time.

Supplementary Table 1, footnote b: “This five-year abstinence period was based on data showing mortality risks by years since smoking cessation. [Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013] We performed an analysis in which we assumed that mortality risks decreased to “Former Smoker” levels immediately upon quitting smoking, but the model-generated results were not a better fit to NHIS data (results not shown).”

18. Supplement - Mortality Inputs: The NHIS Linked mortality follow-up only collects smoking behavior information at baseline, and not at follow-up. This means that someone categorized as a current smoker at baseline, may well have been a former smoker by time of death. Did the authors perform any type of right censoring to avoid the potential for misclassification?

Response: We did not perform right censoring, and therefore there may have been some misclassification. However, the impact of this is expected to be small as we did not consider mortality to decrease until five years after cessation. We have modified this section of the Supplement.

Supplement, page 4: “Because smoking behavior data were collected only at baseline in NHIS and not again at the time of death, there may have been some misclassification of smoking status (e.g., someone who was a current smoker at the time of NHIS assessment may have subsequently quit and later died but was still considered a current smoker). However, all-cause mortality rates do not significantly decrease until a few years after cessation, and we considered those who had quit smoking to have similar mortality risks to current smokers until five years of abstinence.” (Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013)

Reviewer 2

19. Major: Figure 3 and external mortality validation: It is important to additionally validate against ex-smoker mortality (or overall mortality). The inclusion of relapse times and the calibrated cessation probabilities have biggest impact on ex-smokers’ mortality and this crucial sub-population needs to be included in the external mortality validation preferably directly or indirectly by using overall mortality not stratified by status (that includes never, ever, and current smokers).

Response: We recognize the Reviewer’s point about the importance of examining the outcomes of former smokers. Their age- and sex-specific mortality rates in the external validation are shown in Supplementary Table 2. In this table, we were able to show the mortality rates because NHIS data on years since quitting were available (Supplementary Text, page 3). However, we deliberately did not include former smokers in the cumulative mortality curves in Figure 3 because these curves assume a consistent smoking status over time, and age at cessation – a critical determinant of mortality risk – would not be captured. The vertical axis in Figure 3 is “cumulative mortality from age 20.” We depicted those who are current smokers at age 20 and continue to smoke until death, and those who are never smokers at age 20 and do not subsequently start smoking. It is not informative to include those who are already former smokers at age 20: if they remain abstinent from smoking, their mortality risks are equal to those of never smokers (Jha et al., N Engl J Med 2013); if they relapse to smoking, then they become current smokers again. We have modified the Methods for clarification.

Methods, page 16: “We also produced curves of cumulative mortality from STOP-generated results and from NHIS data, stratified by sex and by current/never smoking status. These curves reflect 20-

year-old current smokers who continue to smoke until death or 20-year-old never smokers who never start smoking. We compared the four sets of cumulative mortality curves by the RMSE and CV-RMSE (STOP versus NHIS) from age 20 years until age 84 years (goal RMSE <0.01). We did not generate mortality curves for 20-year-old former smokers because mortality risks for those who stop smoking prior to age 20 are similar to those of never smokers. [Jha et al., N Engl J Med 2013] Also, mortality risks depend on age at cessation, and at older ages this heterogeneous group would include people who quit smoking at a variety of ages. [Jha et al., N Engl J Med 2013; Thun et al., N Engl J Med 2013]"

20. P18 last paragraph (cont. in p19) it completely ignores calibration drift, a well-documented phenomenon. My understanding is that the current implementation of STOP does nothing to prevent calibration drift, therefore, the model is unsuitable "to predict future tobacco use" as the authors claim.

Response: We agree that model projections of future tobacco use and associated morbidity and mortality outcomes are subject to uncertainty in the input parameters that govern them. To account for this, we will adhere to the guidelines of the International Society for Pharmacoeconomics and Outcomes Research – Society for Medical Decision Making (ISPOR-SMDM) Modeling Good Research Practices Task Force (Briggs et al., Med Decis Making 2012), which recommend deterministic and probabilistic sensitivity analysis when using the model for forecasting or for evaluating interventions. We also acknowledge the potential challenges in predicting future tobacco use. We have modified the Discussion:

Discussion, page 23: "Going forward, we plan to use the STOP model to study contemporary rather than historical populations and to predict future tobacco use, while utilizing deterministic and probabilistic sensitivity analyses to account for uncertainty in future behavioral transition probabilities and mortality probabilities. [Briggs et al., Med Decis Making 2012]"

Discussion, page 25: "The STOP model has limitations. Its projections are limited by assumptions and the degree of specificity of available data – for example, age, sex, and birth year stratifications of smoking behavioral transitions. While we have aimed to calibrate and validate the model with the best available historical data, any use of the model to project future outcomes should be approached with prudence. Calibration on historical data is no panacea because of concerns of calibration drift."

21. Please discuss the fact that calibration of cessation rates may (over)compensate for other inaccuracies in model inputs or model structure. Your mentioned sensitivity analysis where you simultaneously calibrated initiation rates and the alignment of your multiplier with empirical evidence partially addresses the issue.

Response: We recognize that all models are susceptible to structural misspecifications (George Box: "All models are wrong, but some are useful") and that conclusions from any model are conditional on its structure. That said, we have now mentioned this limitation in the Discussion.

Discussion, page 22: "Calibration of cessation rates may compensate for other inaccuracies in model inputs or structure, though the pre-calibration (without relapse) STOP-generated results fit well with those of CISNET."

22. Calibration on historic data or other models is not panacea, and while the potential future uses of the model are enthusiastically discussed, the limitations and implicit assumptions they involve are understated.

Response: We have now stated this in the Discussion.

Discussion, page 25: “The STOP model has limitations. Its projections are limited by assumptions and the degree of specificity of available data – for example, age, sex, and birth year stratifications of smoking behavioral transitions. While we have aimed to calibrate and validate the model with the best available historical data, any use of the model to project future outcomes should be approached with prudence. Calibration on historical data is no panacea because of concerns of calibration drift.”

23. Minor: The introduction is not completely relevant to the main body of the paper which is methodological in nature. There are themes like e-cigs utilisation or that “the decline has not been seen in all segments of society” that are only remotely relevant if at all to the presented research.

Response: We agree that specific mention of electronic cigarettes may detract a reader from the methodologic aims of this paper, which do not involve electronic cigarettes. We have modified the Introduction, including combining the first two paragraphs into a single paragraph.

Introduction, page 5: “Meanwhile, tobacco treatment interventions, including behavioral therapy and pharmacotherapy, remain underutilized. Novel tobacco and nicotine products, including electronic cigarettes (e-cigs) and heated tobacco products, raise many new clinical and policy questions.”

We agree that the main body of the paper is methodologic in nature and that some statements in the Introduction are not directly related to methods. However, we believe that it is important to convey why novel tobacco research remains critical despite a decrease in the overall prevalence of cigarette smoking in the US, and why model-based research is particularly important given that many traditional methods of clinical research cannot be used to address questions around novel tobacco and nicotine products in a timely manner. Thus, our Introduction is intended to provide a clinical and public health rationale for developing a new model.

24. Please provide a brief description of the CISNET model which you use to calibrate STOP (or move the one from the supplement).

Response: We have moved the description of the CISNET model from the Supplement to the main text.

Methods, page 11: “CISNET is a collaboration of National Cancer Institute-supported investigators modeling the impact of interventions on population incidence and mortality of various types of cancer, including lung cancer. The Yale CISNET-Lung models, for subsequent analyses of cancer care interventions, used data from NHIS to generate detailed smoking initiation and cessation rates, stratified by birth year, age, and sex, and mortality rates, stratified by birth year, age, sex, and smoking status.”

25. It is not entirely clear to me why you chose the term internal validation instead of cross validation for your comparison with the CISNET model. Do they share common inputs?

Response: While there are inconsistent and sometimes contradictory definitions of validity of microsimulation models in the literature (Kopec et al., BMC Public Health 2010), we agree with the Reviewer that our comparison with the CISNET model can be considered cross-validation. Therefore, we have made the necessary changes throughout the manuscript. For example:

Methods, page 10:
"Cross-validation
Overview and outcome comparisons
We conducted cross-validation by simulating the US population born in 1950, following them monthly until 2020, and then comparing STOP results to those from CISNET modeling studies…"

26. I find table 2 very confusing to the extent that I cannot offer a suggestion for improvement. Splitting
the table into 3 tables (one for internal and 2 for external validation) may be a good starting point.

Response: We have made several modifications to Table 2. We created separate row headers for those parameters that were derived from CISNET/NHIS and those that were derived from smoking studies or were calibrated. Where the same parameter was applied in different validation exercises, we indicated the parameter in each column instead of spreading it across multiple columns in a merged cell. Because of the journal's limitations on number of tables and figures, we have not split the single table into 3 tables.

Table 2: See multiple modifications.

27. P12 L6. Other microsimulations in the past have included relapse probability in their calculations. Please tune down the phrase "A novel aspect of the STOP model".

Response: We have modified the phrase in the Methods and the Discussion, while acknowledging the novelty of a monthly time cycle as mentioned by Reviewer 2.

Methods, page 15: "The STOP model specifically includes smoking relapse…"

Discussion, page 22: "A novel aspect of the STOP model is the incorporation of smoking relapse on a monthly basis, reflecting the understanding of nicotine addiction as a chronic relapsing condition with rapid cycles between use and cessation."

28. METHODS: Authors clearly described definitions of transient quit attempts and sustained abstinence. However, they were only mentioned in the middle part of method section. Authors could define those terminologies in the first few paragraphs to give readers a clear idea.

Response: We have moved the paragraph describing transient quit attempts and sustained abstinence to an earlier part of the Methods, within a "Smoking definitions" subsection.

Methods, page 9: "To better distinguish relapse and mortality risks among those with short-term or long-term abstinence, the STOP model includes three states…"

29. P18 L30-46. The approach that is described here to assess potential efficacy of interventions requires
strong, not data-driven assumptions of the modelled intervention on reducing relapse probability in the long run. The way it is described it compromises the usefulness of the model in assessing efficacy.

Response: We have modified this section of the Discussion to more clearly describe the utility of a model analysis in combining short-term trial results with those of longer-term studies of the natural history of smoking and smoking cessation.

Discussion, page 22: “Many trials of smoking cessation interventions follow patients for a few months or up to one year, but they do not report subsequent relapse. By including relapse, the STOP model can combine data from short-term trials of smoking cessation interventions with data from natural history studies of smoking and smoking cessation to project longer-term outcomes including sustained abstinence. In capturing changes in an individual’s smoking behaviors over time, the STOP model can assess the efficacy of tobacco cessation interventions both in the short-term, by the interventions promoting quit attempts, and in the long-term, by the interventions reducing relapse and promoting sustained abstinence. The flexibility to integrate data from a variety of sources is a strength of modeling analyses.

VERSION 2 – REVIEW

REVIEWER	Jamie Tam Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, United States One of the co-authors, A David Paltiel, is a colleague in my department at the Yale School of Public Health, which I only recently joined. We have no history of co-authorship or research collaboration.
REVIEW RETURNED	27-Feb-2020

GENERAL COMMENTS	The authors have adequately addressed reviewers' comments. Minor notes: "Tobacco policymakers must consider how emerging products will change cigarette smoking behaviors and clinical outcomes." This sentence in the abstract does not seem to fit with the rest of the paper, which largely ignores discussion of emerging products and seems to pitch the model as more suited for tobacco cessation efforts (given its focus on relapse). Instead, the opening sentence of the abstract should describe how relapse is commonplace, despite the absence of simulation models that account for this. Because the authors are calibrating to older data (pre-2009), there should be some acknowledgement in the discussion section that relapse rates could be changing over time, as levels of nicotine dependence in the remaining smoking population may be different today than in the past. Perhaps there is evidence on trends in smoking cessation relapse that could be cited.

REVIEWER	Chris Kypridemos

	University of Liverpool, UK
REVIEW RETURNED	17-Feb-2020

GENERAL COMMENTS	The authors have addressed all my comments & suggestions adequately in their revised version of their manuscript. Therefore, I am happy to recommend their manuscript for publication.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

1. "Tobacco policymakers must consider how emerging products will change cigarette smoking behaviors and clinical outcomes." This sentence in the abstract does not seem to fit with the rest of the paper, which largely ignores discussion of emerging products and seems to pitch the model as more suited for tobacco cessation efforts (given its focus on relapse). Instead, the opening sentence of the abstract should describe how relapse is commonplace, despite the absence of simulation models that account for this.

Response: We have modified the Abstract as suggested.

Abstract, page 3: “Simulation models can project effects of tobacco use and cessation and inform tobacco control policies. Most existing tobacco models do not explicitly include relapse, a key component of the natural history of tobacco use. Our objective was to develop, calibrate, and validate a novel individual-level microsimulation model that would explicitly include smoking relapse and project cigarette smoking behaviors and associated mortality risks.”

2. Because the authors are calibrating to older data (pre-2009), there should be some acknowledgement in the discussion section that relapse rates could be changing over time, as levels of nicotine dependence in the remaining smoking population may be different today than in the past. Perhaps there is evidence on trends in smoking cessation relapse that could be cited..

Response: We have now acknowledged this in the Discussion and cited a paper describing trends in smoking cessation and relapse.

Discussion, page 24: “Calibration on historical data is no panacea because of concerns of calibration drift, and relapse rates could change over time due to changes in population-level nicotine dependence [Yi et al., Tob Induc Dis 2017].”