PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Adolescent Stress Experiences over Time Study (ASETS) Protocol: Design and Methods of A Prospective Longitudinal Study of Sexual Minority Adolescents in the United States
AUTHORS	Schrager, Sheree; Mamey, Mary Rose; Rhoades, Harmony; Goldbach, Jeremy

VERSION 1 – REVIEW

REVIEWER	Mike Parent
REVIEW RETURNED	02-Aug-2021

GENERAL COMMENTS	This is a review of bmjopen-2021-054792, “The Adolescent Stress Experiences over Time Study (ASETS):A Prospective Longitudinal Study of Sexual Minority Adolescents in the United States,” a protocol submitted to BMJ Open. This protocol appears to be clear in terms of background, methodology, and analytic approach. I have only a few comments. 1. The project relies heavily on the SMASI. This is fine in general, but the SMASI appears to me to be a very broad measure that may risk some low fidelity in the data. For example, the work subscale contains items about being physically assaulted at work and being threatened, as well as it just generally being “hard” to be at the workplace due to LGBTQ identity. I do wonder if specific assessment of the most impactful forms of discrimination in an actuarial sense (being assaulted, being kicked out of the home) might be important in addition to the SMASI. 2. In terms of ethics, the protocol for managing positive responses for suicide items needs to be explained; it does not seem ethical to this reader to collect this data and then do nothing for those who endorse suicidal ideation while part of the study. 3. Outness variables on page 23: How does a respondent answer if they are partially out to those in the category (e.g., to some teachers)?

REVIEWER	Allen Mallory The Ohio State University
REVIEW RETURNED	28-Sep-2021

GENERAL COMMENTS	This study describes the protocol used to recruit a national sample of sexual minority adolescents (SMA) and collect follow up surveys six times over three years. Overall, was excited about the protocol and the data that the study will produce. The study will address many of the important gaps the authors have identified with studies

	examining minority stress and health among SMA. The methodological approach is rigorous and generally well described. My first main critique is that the authors understate the current body of literature related to research on SMA in the introduction of their protocol. I note several examples below, but I think this will be relevant for the authors in this paper as well as when they move towards publishing from their data to be able to articulate the contribution made given existing research. Many of the citations used to make some of the over statements were outdated in the sense that more recent research has been done that addresses some of the critiques The authors may need to revisit the literature to update some of their bigger statements early on. My second main critique was that the analyses could be written more clearly—the analyses are complex, so clarity would aid in understanding the planned analyses for the project. I provide more specific feedback below. Introduction • In the intro to the abstract that authors say "...no study has comprehensively examined how minority stress may change throughout the course of adolescence, nor how stress trajectories may predict differences in health during this critical developmental period." Predict should be removed since the authors are not predicting differences, they are just predicting health among SMA. • "The longitudinal cohort design permits the first examination of change in minority stress experiences over time? Should be edited to say, "among adolescents". There are a few other longitudinal studies of adults. • The following paragraph is generally true, but the citations are old and there has been more research done looking at race and ethnic differences among SMY (including SMA) in the last 10 years. See below for a few examples. I also noted similar issues in point 3 of the gaps in the literature. o "While there are likely to be subgroup differences among racial and ethnic lines as well, scholars have noted a relative absence of racial and ethnic diversity in sexual minority research, (15-17) and even in large meta-analytic studies, the lack of racial and ethnic diversity in sampling is noted as a significant limitation.(12, 18) Understanding the experience of these youth is increasingly relevant, as recent national survey data suggest that racial and ethnic minority youth are more likely than white segments of the U.S. population to identify as SMA.(19)" o Toomey, R. B., Huynh, V. W., Jones, S. K., Lee, S., & Revels-Macalinao, M. (2017). Sexual minority youth of color: A content analysis and critical review of the literature. Journal of Gay & Lesbian Mental Health, 21(1), 3–31. https://doi.org/10.1080/19359705.2016.1217499 o Layland, E. K., Exten, C., Mallory, A. B., Williams, N. D., & Fish, J. N. (2020). Suicide Attempt Rates and Associations with Discrimination Are Greatest in Early Adulthood for Sexual Minority Adults Across Diverse Racial and Ethnic Groups. LGBT Health, 7(8), 439–447. https://doi.org/10.1089/lgbt.2020.0142 o Mallory, A. B., & Russell, S. T. (2021). Intersections of Racial Discrimination and LGB Victimization for Mental Health: A Prospective Study of Sexual Minority Youth of Color. Journal of Youth and Adolescence, 50(7), 1353–1368. https://doi.org/10.1007/s10964-021-01443-x o Bostwick, W. B., Meyer, I., Aranda, F., Russell, S., Hughes, T., Birkett, M., & Mustanski, B. (2014). Mental Health and Suicidality Among Racially/Ethnically Diverse Sexual Minority Youths.

	American Journal of Public Health, 104(6), 1129–1136. https://doi.org/10.2105/AJPH.2013.301749 o Thoma, B. C., & Huebner, D. M. (2013). Health consequences of racist and antigay discrimination for multiple minority adolescents. Cultural Diversity and Ethnic Minority Psychology, 19(4), 404–413. https://doi.org/10.1037/a0031739 o Pollitt, A. M., Mallory, A. B., & Fish, J. N. (2018). Homophobic Bullying and Sexual Minority Youth Alcohol Use: Do Sex and Race/Ethnicity Matter? LGBT Health, 5(7), 412–420. https://doi.org/10.1089/lgbt.2018.0031 • Also in the second paragraph of the introduction the authors make a lot of arguments for gaps in the literature for SMA, specifically, rather than SMY, which often include young adults 18-24. This focus on SMA would mean that currently, we do not know what the racial-ethnic diversity of SMA looks like (though we suspect it to be high) so citing the Gates study for this doesn't align with the focus on SMA (defined as 14-17 by authors) since it was of adults 18+ (though it would include SMY). • Page 8 line 31 the authors state: "However no study has ever comprehensively examined this relationship" – could you say what you mean by comprehensively? There are certainly several studies that have looked longitudinally at some of the things you mention above line 31, but maybe not in the way that is later proposed, so this should be more excited. One notable example is the Birkett et al., 2015 study you all cite. • For point 3, of the gaps the authors are accurate that there is limited research on subgroups of SMA, but I think they overstate the paucity of research on subgroups of SMA (See some of citations above). There had been more than just a handful of studies on adults in these areas and the authors should acknowledge this. • Two gaps in the literature that the authors might consider addressing regarding existing longitudinal research with SMA, and that their study may address is the typically poor follow up rates in existing longitudinal studies of SMA. And, though the authors propose an approach that sounds promising, could they elaborate more on retention efforts? • Another gap that might be mentioned (if the authors try to address it) is how even when studies of SMA are large, they often still have small samples of the subgroups the authors want to explore—it would be in helpful to mention and discuss this in the protocol if it is addressed. There is some concern that the subgroup comparisons such as race ethnicity may end up being just being white vs non-white, which only addresses existing limitations in a cursory way—the same could be said for urbanicity. • In the current study section page 11 line 10, point (d) was a surprise as it is not mentioned earlier as a limitation. The authors definition of SMA in the introduction does not match their definition in the methods (i.e., identity vs attraction or identity). Also, several other studies, of adolescents in particular use multiple indicators of sexual orientation as SMA they may not be out to people yet. The authors should be clear about what this "method for identifying hidden populations" they implemented was so it's clear what the contribution is beyond what is already done.

	Method • Under population definition, page 11 line 26-29 current data suggests that closer to 15% of SMA adolescents/ young adults identify as a sexual minority using Gallup data and YRBS data: o https://news.gallup.com/poll/329708/lgbt-identification-rises-latest-estimate.aspx o Raifman, J., Charlton, B. M., Arrington-Sanders, R., Chan, P. A., Rusley, J., Mayer, K. H., Stein, M. D., Austin, S. B., & McConnell, M. (2020). Sexual Orientation and Suicide Attempt Disparities Among US Adolescents: 2009–2017. Pediatrics, 145(3). https://doi.org/10.1542/peds.2019-1658 • Why not included sexual behavior as well as attraction and identity to conceptualize sexual orientation? • A justification is needed for excluding trans and non-binary adolescents as we know increasingly youth identify in these ways (see below) and many also identify as a SM. Also given the longitudinal nature of the study during a critical developmental period it would not be surprising to see some shifts in SMA gender identity over time. o Chew, D., Tollit, M. A., Poulakis, Z., Zwickl, S., Cheung, A. S., & Pang, K. C. (2020). Youths with a non-binary gender identity: A review of their sociodemographic and clinical profile. The Lancet Child & Adolescent Health, 4(4), 322–330. https://doi.org/10.1016/S2352-4642(19)30403-1 • Could authors make clearer which questions/items are planned for longitudinal follow up? Analysis • In general, this section was a bit hard to follow, so one suggestion is to make sure that the same terms are being used to describe the analyses throughout the analysis section. • I found myself wondering if the authors have enough power to do many of the proposed analyses, even just the primary cohort sequential LGCM. Will there be enough SMA in each bucket/cohort? (14, 15, 16, and 17) especially considering attrition? Based on the power analysis, the sample may already be underpowered. Do the authors have alternative analyses plans given these concerns? • Could the authors explain more what they mean by “Proper functional forms of trajectories will be identified prior to estimating full unconditional LGCMs.(80)”? This is usually the first step in conducting a LGCM (i.e., the unconditional model). Unless the authors mean that they will first look at the sample mean scores before estimating an unconditional LGCM? Even so, it’s still best practice to test which form best fits the data (using model fit and chi-square tests as described) even if sample means look a certain way. • A related point, how will the binary indicators be incorporated into the LGCM? Models with suicide items and substance use variables would need additional consideration given that they would not be normally distributed. • It is unclear what is meant by “We will test piecewise models with one (i.e., linear across development) to four growth trajectories to

	determine the number of different stages of adolescence that best represent the data." Yes, you will have an appropriate number of waves to do this, but only middle adolescence is really captured in the sample (with participants growing into young adulthood/late adolescence). So how would multiple stages of adolescence be captured? Is it really just age differences or two development periods: middle adolescence and young adulthood/late adolescence? • For WH2.1, the authors could be clearer about how the parallel process models will be specified. Will direction hypotheses be specified in only one direction (i.e., minority stress predicting health) or will directional paths be investigated from the outcome to minority stress as well? • Also, how will the authors estimate both cohort sequential and piecewise models? This needs to be more clearly explained. My understanding of the cohort sequential model is that it is essentially a combination of a multiple group model (each age as a group) and a piecewise model (which connects the slopes of each age group to creates a LGCM capturing changes with age). So, its not clear how there could be a cohort sequential and (within that) piece wise LGCM (especially if you plan to have a similar kind of model for the outcome variables). • For WH3.1, could the authors specify how many groups will be considered for stratification variables? • Again, I will raise concerns regarding power with the complexity of models proposed. For the multiple group models, my understanding is that a cohort sequential parallel process model with be run within two or more groups, which might be fine for two groups, but more problematic as with more than two groups, which you might for race/ethnicity and sexual identity. • It might also be worth while testing the group and longitudinal invariance of each scale -- this would also be a contribution to the field, but also important give the argument that these are novel analyses it would be important to ensure the measure work the same across diverse groups.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 (Dr. Mike Parent)
Comments to the Author:

This is a review of bmjopen-2021-054792, "The Adolescent Stress Experiences over Time Study (ASETS):A Prospective Longitudinal Study of Sexual Minority Adolescents in the United States," a protocol submitted to BMJ Open.

This protocol appears to be clear in terms of background, methodology, and analytic approach. I have only a few comments.

1. The project relies heavily on the SMASI. This is fine in general, but the SMASI appears to me to be a very broad measure that may risk some low fidelity in the data. For example, the work subscale contains items about being physically assaulted at work and being threatened, as well as it just generally being "hard" to be at the workplace due to LGBTQ identity. I do wonder if specific assessment of the most impactful forms of discrimination in an actuarial sense (being assaulted, being kicked out of the home) might be important in addition to the SMASI.

We absolutely agree that the proposed analyses rely heavily on the functioning of the SMASI, as one of the primary purposes of this study was to extend the original validation of the measure (which was developed and preliminarily validated on a smaller, mostly regional sample) to a large national sample of SMA and verify its measurement invariance over time. Although the measure is "very broad" in the sense that its subscales cover multiple domains, this is intentional – the SMASI was designed for flexible use, so that it could be modeled as a total score of the 54 main items, a latent variable with manifest subscale scores, or the subscales could be modeled independently. As noted by the reviewer and described in the development of the measure (Schrager et al. 2018), the full measure does contain a mix of cognitive appraisals, emotion-based questions, and explicit objective and time-limited experiences (e.g., physical assault), and the whole-scale and subscale item sets do consistently cohere. Of note, the work subscale is unique in that it is only assessed among adolescents who have an employment history, i.e. are currently working or have previously held a job. Reliability of the work subscale has typically been high (ω=.96, as noted on p. 20), but as adolescents get older over the course of the current study and more of them enter the workforce, we will be monitoring this subscale to ensure that it continues to perform well.

2. In terms of ethics, the protocol for managing positive responses for suicide items needs to be explained; it does not seem ethical to this reader to collect this data and then do nothing for those who endorse suicidal ideation while part of the study.

We thank the reviewer for this important observation and fully agree about the ethical imperative to attend to study participants who may be experiencing suicidality. Our approach to monitoring and addressing suicidality has been further explained in the section on ethics and dissemination (pp. 32-33), including referral to available resources: "Because the study is purely observational with no researcher-controlled intervention, there is no external data safety monitoring board for the study. However, a member of the research team reviews study data immediately upon downloading the new data files each business day, and any open-ended statements or data that could potentially suggest participant safety concerns are immediately brought to the attention of the study investigators, who are considered mandated reporters in the State of California. Statements are reviewed and assessed for information concerning abuse or neglect of a child; abuse or neglect of an elder; or threat that the participant will harm themselves or someone else. An IRB-approved standard operating procedure is in place in the event of a positive disclosure; however, to date, no participant has disclosed any imminent safety concerns, and no other adverse events have been reported. The protocol included providing referrals to support resources for all participants, and following up with specific additional resources for those who screened at risk for suicidality (e.g., both general and LGBTQ-specific crisis services)."

3. Outness variables on page 23: How does a respondent answer if they are partially out to those in the category (e.g., to some teachers)?

We have reworded the description of this variable to clarify that the disclosure items are binary items, and youth would answer "yes" if they had disclosed their sexual identity to anyone in that category (e.g., at least one teacher), "no" if they had not disclosed to anyone in that category, and "N/A" if the entire category were inapplicable to them (e.g., participant does not have siblings).

Reviewer: 2
Dr. Allen Mallory, The Ohio State University
Comments to the Author:
This study describes the protocol used to recruit a national sample of sexual minority adolescents (SMA) and collect follow up surveys six times over three years. Overall, was excited about the protocol and the data that the study will produce. The study will address many of the important gaps the authors have identified with studies examining minority stress and health among SMA. The methodological approach is rigorous and generally well described. My first main critique is that the authors understate the current body of literature related to research on SMA in the introduction of their protocol. I note several examples below, but I think this will be relevant for the authors in this paper as well as when they move towards publishing from their data to be able to articulate the contribution made given existing research. Many of the citations used to make some of the over statements were outdated in the sense that more recent research has been done that addresses some of the critiques The authors may need to revisit the literature to update some of their bigger statements early on. My second main critique was that the analyses could be written more clearly—the analyses are complex,

so clarity would aid in understanding the planned analyses for the project. I provide more specific feedback below.

We thank the reviewer for this feedback and have endeavored to integrate these suggestions (specifics below) and reframe the manuscript in the context of the updated literature review.

Introduction

- *In the intro to the abstract that authors say "...no study has comprehensively examined how minority stress may change throughout the course of adolescence, nor how stress trajectories may predict differences in health during this critical developmental period." Predict should be removed since the authors are not predicting differences, they are just predicting health among SMA.*

We thank the reviewer for the suggestion to be more specific about our outcome of study. Because this is a longitudinal study with planned causal analyses using eventual distal (3-year) outcome measures, we felt the "predictive" language is important and have reworded the sentence to read "predict health outcomes during this critical developmental period" rather than "differences in health outcomes."

- *"The longitudinal cohort design permits the first examination of change in minority stress experiences over time? Should be edited to say, "among adolescents". There are a few other longitudinal studies of adults.*

We thank the reviewer for this recommendation and have made this change in the Strengths and Limitations section of the manuscript.

- *The following paragraph is generally true, but the citations are old and there has been more research done looking at race and ethnic differences among SMY (including SMA) in the last 10 years. See below for a few examples. I also noted similar issues in point 3 of the gaps in the literature.*
  - *"While there are likely to be subgroup differences among racial and ethnic lines as well, scholars have noted a relative absence of racial and ethnic diversity in sexual minority research, (15-17) and even in large meta-analytic studies, the lack of racial and ethnic diversity in sampling is noted as a significant limitation.(12, 18) Understanding the experience of these youth is increasingly relevant, as recent national survey data suggest that racial and ethnic minority youth are more likely than white segments of the U.S. population to identify as SMA.(19)"*

We greatly appreciate the recommendation to provide a more comprehensive and accurate picture of the current state of the literature on sexual minority youth of color. We referenced two additional studies and softened the language of this paragraph (pp. 5-6) to focus on the need for additional intersectional work; further on, in the more detailed section on subgroup differences (pp. 7-8), we incorporated the recommended citations to the manuscripts by Toomey et al., Layland et al., Mallory et al., Bostwick et al, Thoma & Huebner, and Pollitt et al., and revised this section to focus on the areas where our study will make the greatest additional contribution, i.e. by characterizing the determinants of subgroup health disparities and exploring them over the course of adolescence.

- *Also in the second paragraph of the introduction the authors make a lot of arguments for gaps in the literature for SMA, specifically, rather than SMY, which often include young adults 18-24. This focus on SMA would mean that currently, we do not know what the racial-ethnic diversity of SMA looks like (though we suspect it to be high) so citing the Gates study for this doesn't align with the focus on SMA (defined as 14-17 by authors) since it was of adults 18+ (though it would include SMY).*

We apologize for the imprecision in the previous draft, and have revised this part of the second paragraph (p. 6) to more clearly delineate the data on adolescents vs. youth, as follows: "Recent studies of Black and Latinx adolescents do suggest that intersectional differences may exist, and understanding their experience is increasingly relevant: although population estimates specific to adolescents are lacking, national survey data suggest that racial and ethnic minority youth are more likely than white segments of the U.S. population to identify as SMA."

- *Page 8 line 31 the authors state: "However no study has ever comprehensively examined this relationship" – could you say what you mean by comprehensively? There are certainly several studies that have looked longitudinally at some of the things you mention above line 31, but maybe*

not in the way that is later proposed, so this should be more excited. One notable example is the Birkett et al., 2015 study you all cite.

We have edited this sentence (now on p. 6 lines 18-20) to refer specifically to adolescents, in keeping with the change made earlier in the introduction, and to clarify that we mean "comprehensively examined the relationship between minority stress and health outcomes longitudinally among adolescents." That is, individual studies have assessed experiences that we might categorize as individual minority stressors, but the main gap has largely been with the measurement of minority stress, which could not be reliably and validly measured for adolescents until 2018. Without that measure, a comprehensive understanding of the interplay between various indicators of the multifaceted minority stress construct and behavioral health outcomes has not been possible.

- *For point 3, of the gaps the authors are accurate that there is limited research on subgroups of SMA, but I think they overstate the paucity of research on subgroups of SMA (See some of citations above). There had been more than just a handful of studies on adults in these areas and the authors should acknowledge this.*

We appreciate these remarks and have revised the manuscript throughout, as described in more detail above, to acknowledge the developments in the literature since this study was originally proposed and focus on the lack of longitudinal studies of diverse sexual minority adolescents where minority stress is also measured well.

- *Two gaps in the literature that the authors might consider addressing regarding existing longitudinal research with SMA, and that their study may address is the typically poor follow up rates in existing longitudinal studies of SMA. And, though the authors propose an approach that sounds promising, could they elaborate more on retention efforts?*

We thank the reviewer for this suggestion and have added a sentence about the importance of establishing feasibility for retention of SMA in prospective studies in the paragraph describing the lack of studies over time (p. 7, lines 19-20). Additionally, our retention approaches are outlined in the methods, with more detail provided to describe both the monthly "check-ins" and post-survey data collection (pp. 17-18).

- *Another gap that might be mentioned (if the authors try to address it) is how even when studies of SMA are large, they often still have small samples of the subgroups the authors want to explore—it would be in helpful to mention and discuss this in the protocol if it is addressed. There is some concern that the subgroup comparisons such as race ethnicity may end up being just being white vs non-white, which only addresses existing limitations in a cursory way—the same could be said for urbanicity.*

We appreciate and agree with this limitation of large studies of SMA, and had noted it in the earlier literature review on p. 6 (i.e., "Even in large meta-analytic studies that include young adults, the lack of racial and ethnic diversity in sampling is noted as a significant limitation.") From this suggestion, we have added in the gap section describing subgroup differences (p. 8) the following sentence: "A large study that is well-powered to examine differences among multiple demographic groups – that is, able to model more than simple binary comparisons – remains needed."

- *In the current study section page 11 line 10, point (d) was a surprise as it is not mentioned earlier as a limitation. The authors definition of SMA in the introduction does not match their definition in the methods (i.e., identity vs attraction or identity). Also, several other studies, of adolescents in particular use multiple indicators of sexual orientation as SMA they may not be out to people yet. The authors should be clear about what this "method for identifying hidden populations" they implemented was so it's clear what the contribution is beyond what is already done.*

We apologize for the lack of clarity in this point and have revised the sentence to specify only that we were referring to using respondent-driven sampling (RDS) to complement our direct advertising methods. RDS has been used successfully with adult "hidden" populations, but only rarely with youth samples. The revised sentence (p. 9, lines 16-18) now reads, "(d) respondent-driven sampling methods to recruit youth who may have not disclosed their sexual orientation to others, including their parents…"

Method

- *Under population definition, page 11 line 26-29 current data suggests that closer to 15% of SMA adolescents/ young adults identify as a sexual minority using Gallup data and YRBS data:*

We thank the reviewer for the newer references and have incorporated these updated population estimates on page 10.

- *Why not included sexual behavior as well as attraction and identity to conceptualize sexual orientation?*

The question of how to categorize sexual identity for the purpose of determining study eligibility is complex. There is a practical limitation based on the fact that this was a one-step, internet-based survey – we had to determine very quickly, with closed-ended items, whether a potential participant was eligible or not in order to use logic built into the Qualtrics platform to automatically categorize someone as eligible (in which case they were taken to the assent form) or ineligible (in which case the survey ended). This precluded us from allowing more nuanced and complete measures, such as write-in (free-text) sexual identity labels, to be used to determine eligibility. There is also an ethical limitation based on the fact that our IRB permitted us to use benign demographic screening items for the eligibility determination, before eligible participants were asked to provide informed assent. While basic questions about attraction and identity were appropriate to include as a screener, questions about sexual behavior were deemed too sensitive to include in a screening tool prior to assent.

Given that the theoretical framework underpinning the study – minority stress theory – is at its heart a theory of identity, we were comfortable moving forward on the broadest possible basis of allowing any youth who identified as not 100% heterosexual to be included provided they met the other inclusion criteria, with the understanding that their specific identity and even attraction would likely change over time. As noted in the *Additional Measures* section (p. 24), we did measure sexual behavior later in the survey, to provide a more complete picture of sexual orientation, and we plan to conduct an in-depth analysis of the concordance (or discordance) between the myriad assessments of identity, attraction, and behavior in our sample. Nonetheless, generalizability based on study eligibility criteria is acknowledged as a limitation of our study.

- *A justification is needed for excluding trans and non-binary adolescents as we know increasingly youth identify in these ways (see below) and many also identify as a SM. Also given the longitudinal nature of the study during a critical developmental period it would not be surprising to see some shifts in SMA gender identity over time.*

We appreciate this comment and recognize that our rationale was underspecified in the original draft. We have edited the population definition section (p. 10) to more clearly articulate a) why the original sample was limited to cisgender adolescents, and b) that we do expect to see changes in gender identity in this sample over time, and that we will retain those participants longitudinally. The revised section now reads: “We also required participants to be cisgender, i.e., to express a gender identity congruent with their sex assigned at birth, at the time of recruitment. At the time this study was open to enrollment, the SMASI had only been validated with cisgender adolescents. As the primary purpose of this study was the longitudinal validation of the SMASI instrument, we chose to mirror the inclusion criteria on which the SMASI was initially developed and validated in order to reduce error variance when assessing minority stress attributable to sexual identity, given the present inability to differentiate between sexual and gender minority stress among adolescents who are both sexual and gender minorities. Although transgender and nonbinary youth were excluded from enrolling in the baseline sample, maintaining a cisgender identity over time was not a requirement for continuation in the longitudinal portion of the study; indeed, we expect changes in gender identity over time, and will assess those in later waves (see *Measures).* A separate NIH-funded study (R21HD082813-01A1) is now underway to examine gender minority stress among transgender and nonbinary youth and parse out sexual versus gender minority stress experiences.”

- *Could authors make clearer which questions/items are planned for longitudinal follow up?*

We have added the following statement after the description of all primary measures (i.e., all outcomes, key predictors, covariates, and protective factors), on p. 24 (lines 3-6): “With the exception

of sex assigned at birth and race/ethnicity, which were only captured at baseline, all of the above measures were also collected at each follow-up time point. This includes explicitly inquiring about sexual identity and gender identity at each wave, given the fluidity of these identities during adolescence."

Analysis

- *In general, this section was a bit hard to follow, so one suggestion is to make sure that the same terms are being used to describe the analyses throughout the analysis section.*

We thank the reviewer for this suggestion and have thoroughly revised the analytic plan to ensure that consistent language is used when describing specific analytic techniques and variables.

- *I found myself wondering if the authors have enough power to do many of the proposed analyses, even just the primary cohort sequential LGCM. Will there be enough SMA in each bucket/cohort? (14, 15, 16, and 17) especially considering attrition? Based on the power analysis, the sample may already be underpowered. Do the authors have alternative analyses plans given these concerns?*

We address this in somewhat more detail in response to the reviewer's questions about statistical power and multigroup modeling near the end of the review, but in brief, we will indeed be adequately powered to test all proposed longitudinal models, including those where we model up to four groups simultaneously, if we achieve our proposed sample of at least N=1,075 overall and n=190 participants in all demographic subgroups of interest (including appropriately collapsed/recoded subgroups). Our strategies for ensuring a sufficient sample are a combination of substantial over-recruitment across the board at baseline (knowing not all participants will elect to enroll longitudinally); extensive retention and sample monitoring methods; and statistical techniques that are robust to participants missing a wave of data collection but rejoining the study at a later wave. If truly necessary, we are able to scale back the analytic approach as needed (e.g., limit piecewise models to linear or quadratic rather than higher-order polynomial "pieces;" limit invariance models to three rather than four group comparison, etc.) However, we powered the study assuming the need to implement the most restrictive possible models (i.e., those leaving us with only 7 degrees of freedom). Based on our prior work with SMA – and in particular experience modeling these specific outcomes in sexual minority youth and young adults – we expect to be able to complete the full set of analyses as proposed.

- *Could the authors explain more what they mean by "Proper functional forms of trajectories will be identified prior to estimating full unconditional LGCMs.(80)"? This is usually the first step in conducting a LGCM (i.e., the unconditional model). Unless the authors mean that they will first look at the sample mean scores before estimating an unconditional LGCM? Even so, it's still best practice to test which form best fits the data (using model fit and chi-square tests as described) even if sample means look a certain way.*

Yes, as the reviewer surmised, our plan is to identify the best fitting form by testing linear, quadratic, cubic and piecewise functions on our trajectories. This is indeed usually the first step in conducting a LGCM, and in the effort to outline the steps we plan to take in the paper, we did not want to omit this. For clarity, we have revised this statement on p. 25 (lines 10-12): "Prior to estimating full unconditional LGCMs, we will identify the best-fitting functional forms of trajectories (i.e., linear, quadratic, cubic, and/or piecewise) for each variable."

- *A related point, how will the binary indicators be incorporated into the LGCM? Models with suicide items and substance use variables would need additional consideration given that they would not be normally distributed.*

We agree with this comment. We previously specified that "Distributional properties of all continuous and categorical variables will be evaluated, and we will apply appropriate transformation or robust estimation procedures to correct for non-normally distributed variables." We have now added the statement "(e.g., specifying the WLSMV estimator for binary indicators)" on p. 25 line 4 to more clearly identify the procedure that accounts for non-normal variable distribution. We will be constructing our models following a similar method used in Mamey et al. (2021), which in turn was based on the recommendations of Muthen (2001), and have added these citations accordingly.

- *It is unclear what is meant by "We will test piecewise models with one (i.e., linear across development) to four growth trajectories to determine the number of different stages of adolescence that best represent the data." Yes, you will have an appropriate number of waves to do this, but only middle adolescence is really captured in the sample (with participants growing into young adulthood/late adolescence). So how would multiple stages of adolescence be captured? Is it really just age differences or two development periods: middle adolescence and young adulthood/late adolescence?*

We thank the reviewer for this comment and agree that middle adolescence will contain the most information/data. Indeed, this is part of the design of the sequential cohort model: to capture multiple ages within a shorter amount of time, adolescents between the ages of 14 and 17 were recruited to understand ages 14-20, and the thinner "tails" of the age distribution curve (i.e., at 14/15 and 19/20) are an inherent part of this design. We took care to ensure that there were enough 14-year-olds recruited at baseline to make sure this age group was well represented, without deliberate oversampling (as their data would later turn into 16- and 17-year data in the cohort sequential model, negating any benefits of over-recruitment).

With this in mind, we agree that language referring to this approach as capturing "stages" of adolescence may have been confusing or misleading. We've removed the phrase "different stages of adolescence" and restated our approach as: "We will test piecewise models with one (i.e., linear across development) to four growth trajectories, based on which best represents the data." We have also added another sentence to explain the significance and importance of these growth trajectories: "This will also allow us to understand how and when changes may occur throughout adolescence." (pp. 25-26)

- *For WH2.1, the authors could be clearer about how the parallel process models will be specified. Will direction hypotheses be specified in only one direction (i.e., minority stress predicting health) or will directional paths be investigated from the outcome to minority stress as well?*

We thank the reviewer for this question, as the original framing was unclear. We are explicitly hypothesizing the direction that minority stress predicts health over time, particularly in light of the minority stress measure capturing cumulative lifetime experiences up to each time point (whereas the health measures are assessed relative to specific recent periods of time, e.g. past two weeks, past 30 days, etc.). To clarify this position, we have revised the sentence on p. 27 (lines 22-23) to read: "Regression coefficients reflecting influence of minority stress on each outcome (i.e., regression of health outcome slopes and intercepts onto SMASI slopes and intercepts) will be estimated."

- *Also, how will the authors estimate both cohort sequential and piecewise models? This needs to be more clearly explained. My understanding of the cohort sequential model is that it is essentially a combination of a multiple group model (each age as a group) and a piecewise model (which connects the slopes of each age group to creates a LGCM capturing changes with age). So, its not clear how there could be a cohort sequential and (within that) piece wise LGCM (especially if you plan to have a similar kind of model for the outcome variables).*

The cohort sequential model is a design choice – that is, it is a restructuring of the data so that individuals' repeated measures are stored (and modeled) by their age at that time, rather than by wave of data collection. The piecewise (or polynomial) models refer specifically to the form of the regression equation specifying the relationship between predictor and outcome variables. In that sense, even though the two approaches rely on the common term "models," we are able to employ both modeling techniques simultaneously as they are not in conflict.

- *For WH3.1, could the authors specify how many groups will be considered for stratification variables?*

We are able to model up to four identity groups simultaneously within a single multi-group (stratification) analysis, now noted explicitly on p. 29 line 5. A more thorough response regarding the statistical power of these models is provided below.

• *Again, I will raise concerns regarding power with the complexity of models proposed. For the multiple group models, my understanding is that a cohort sequential parallel process model with be run within two or more groups, which might be fine for two groups, but more problematic as with more than two groups, which you might for race/ethnicity and sexual identity.*

We understand the reviewer's concern, and we acknowledge that the complexity of parallel growth models need more power than other models to properly converge. However, with more than 1,000 participants and seven time points, we will be adequately powered (.80) to run these parallel growth models with multiple groups for up to five groups (with a minimum number of 190 participants per group). As the perfectly even recruitment across five identity groups may be unrealistic, we believe based on prior experience conducting research with SMA that we can successfully recruit and retain a sample sufficient to ensure 190 participants per group for at least four groups at a time. We have thus revised the power analysis section (p. 31) to provide this additional detail, along with a citation to the work by Preacher, Cai, and MacCallum (2007) that provides the method for estimating power for multigroup covariance structure comparisons: "For WH3.1 and WH3.2, which examine differences in minority stress and outcome trajectories by subgroups, statistical power depends on both group size and total sample size, as we cannot assume identical fit of the initial model in all groups. Using the trajectory forms developed for the previous hypotheses, we can support simultaneous trajectory comparisons with at least 190 participants per group; with a planned sample of N=1,075, we would potentially be adequately powered to examine up to five demographic subgroups simultaneously. However, given the low likelihood of perfectly even recruitment across all demographic strata, it may be more feasible to limit models to four groups to ensure adequate group size without overly condensing across meaningful categories." In particular, this should allow us to examine invariance by race/ethnicity and sexual identity with somewhat more nuance, and we are very well powered to include additional, smaller subgroups for these variables in more straightforward regression models.

• *It might also be worth while testing the group and longitudinal invariance of each scale -- this would also be a contribution to the field, but also important give the argument that these are novel analyses it would be important to ensure the measure work the same across diverse groups.*

We thank the reviewer for this suggestion, and we absolutely agree that this would be a unique and valuable contribution to the field. After some discussion, our team ultimately concluded that the study protocol manuscript was intended to faithfully represent our intentions, hypotheses, and planned analyses at the time we developed the study, as a means of scientific accountability; in that light, it feels disingenuous to describe additional investigative areas of current or future pursuit. If the reviewer or editor feels strongly, we are open to revising the section on invariance testing to reflect examination of other measures beyond the SMASI.

VERSION 2 – REVIEW

REVIEWER	Allen Mallory The Ohio State University
REVIEW RETURNED	28-Dec-2021

GENERAL COMMENTS	Thank you for the detailed responses to my comments. All my concerns have been addressed.