A Structured Approach to Evaluating Life Course Hypotheses: Moving Beyond Analyses of Exposed Versus Unexposed in the Omics Context

Yiwen Zhu, Andrew J. Simpkin, Matthew J. Suderman, Alexandre A. Lussier, Esther Walton, Erin C. Dunn[§], and Andrew D.A.C. Smith[§]
[§]Both senior authors contributed equally to this work. Their names appear alphabetically.

Correspondence to Dr. Erin C. Dunn, Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Simches Research Building 6th Floor, Boston, MA 02114, USA (email: edunn2@mgh.harvard.edu; website: www.thedunnlab.com; phone: +1-617-726-9387; fax: 617-726-0830)

Author affiliations: Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts (Yiwen Zhu, Alexandre A. Lussier, Erin C. Dunn); School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland (Andrew J. Simpkin); Medical Research Council Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK (Matthew J. Suderman, Esther Walton); Department of Psychiatry, Harvard Medical School, Boston, Massachusetts (Alexandre A. Lussier, Erin C. Dunn); Department of Psychology, University of Bath, Bath, UK(Esther Walton); Stanley Center for Psychiatric Research, The Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, Massachusetts (Erin C. Dunn); Henry and Alison McCance Center for Brain Health, Massachusetts General Hospital, Boston, Massachusetts (Erin C. Dunn); Applied Statistics Group, University of the West of England, Bristol, UK (Andrew D.A.C. Smith).

Funding: This work was supported by the National Institute of Mental Health of the National Institutes of Health [grant number R01MH113930 awarded to ECD]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. EW was supported by CLOSER, who is funded by the Economic and Social Research Council and the Medical Research Council (grant reference: ES/K000357/1). The funders took no role in the design, execution, analysis or interpretation of the data or in the writing up of the findings.

Conflict of interest: None declared.

Running head: Structured Life Course Modeling in Omics

Abstract

The structured life course modeling approach (SLCMA) is a theory-driven analytic method that empirically compares multiple prespecified life course hypotheses characterizing time-dependent exposure-outcome relationships to determine which theory best fits the observed data. In this study, we performed simulations and empirical analyses to evaluate the performance of the SLCMA when applied to genome-wide DNA methylation (DNAm). Using simulations, we compared five statistical inference tests used with SLCMA (n=700), assessing the family-wise error rate, statistical power, and confidence interval coverage to determine whether inference based on these tests was valid in the presence of substantial multiple testing and small effects, two hallmark challenges of inference from omics data. In the empirical analyses, we evaluated the time-dependent relationship of childhood abuse with genome-wide DNAm (n=703). In simulations, selective inference and max-|*t*|-test performed best: both controlled family-wise error rate and yielded moderate statistical power. Empirical analyses using SLCMA revealed time-dependent effects of childhood abuse on DNAm. Our findings show that SLCMA, applied and interpreted appropriately, can be used in high-throughput settings to examine time-dependent effects underlying exposure-outcome relationships over the life course. We provide recommendations for applying the SLCMA in omics settings and encourage researchers to move beyond analyses of exposed versus unexposed.

Keywords: life course, structured approach, ALSPAC, omics, DNA methylation, post-selection inference

Abbreviations: SLCMA (structured life course modeling approach); FWER (family-wise error rate); ALSPAC (Avon Longitudinal Study of Parents and Children); CpG (cytosine-phosphate-guanine); DNAm (DNA methylation); CI (confidence interval)

Epidemiologists have long been interested in whether and how exposures over the life course affect later health outcomes. Guided by theories developed in life course epidemiology (**Table 1**), researchers are moving beyond simple comparisons of presence versus absence of exposure to characterize time-dependent exposure-outcome relationships (1). Prior work in life course epidemiology has conceptualized timing effects in numerous ways, examining the role of the developmental timing of exposure (sensitive period hypothesis), number of occasions exposed across time (accumulation of risk hypothesis), proximity in time to exposure (recency hypothesis), and change in exposure status across time (mobility hypothesis). Researchers have adopted this life course perspective, uncovering mechanistic insights that advanced many subfields of public health and medicine (2–6). As different life course hypotheses correspond to distinct theories of disease etiology, efforts to formally compare competing hypotheses and identify those best supported by empirical data are needed to guide prevention and intervention planning.

To address the need for systematic comparisons of life course theories, Mishra and colleagues introduced the structured life course modeling approach (SLCMA) (7). The SLCMA allows researchers to compare a set of *a priori*-specified life course theories and use goodness-of-fit criteria to determine which theory is best supported by empirical data. Smith and colleagues later extended this approach with an alternative statistical model selection strategy that uses least angle regression (8), accommodates both binary and continuous exposures (9,10), and improves the accuracy of selecting the correct hypothesis. More recently, Madathil et al. proposed a Bayesian approach to life course modeling that does not perform variable selection, but rather estimates the posterior probability corresponding to each theoretical hypothesis while assessing the relative importance of a series of life course theories (11). Since its inception, the

SLCMA has been applied in a wide range of non-omics epidemiologic studies, including those examining the time-dependent impacts of childhood trauma, physical activity, or socioeconomic position on psychological, metabolic, and disease outcomes (12–18). Compared to other approaches that consider alternative classifications of the exposure, the SLCMA is better positioned to compare competing life course hypotheses simultaneously. By requiring that life course hypotheses are specified *a priori* based on theory, it prevents post-hoc hypothesis generation following exploratory analyses. Moreover, its model selection feature allows a structured assessment of hypotheses without requiring a saturated model.

The growing availability of high-dimensional biological and phenotype data from longitudinal cohort studies has created new opportunities to assess time-varying exposures in epigenomics, transcriptomics, metabolomics, and other omics settings (19–21). While large cross-sectional omics studies have identified *associations* between biological differences and various traits (22), applications of the SLCMA to longitudinal data and high dimensional outcomes allow researchers to answer more complex questions about disease *mechanisms*. For example, Dunn and colleagues applied the SLCMA in a longitudinal birth cohort study to model timing effects of childhood adversity on DNA methylation, which is a widely studied epigenetic mechanism that could give rise to altered gene expression and phenotypic changes. Using the SLCMA, they found that DNAm differences were largely explained by the age at exposure, with the first three years of life appearing to be a sensitive period associated with more DNAm differences. Their results also showed that the SLCMA could identify associations not identified by an epigenome-wide association study of exposed versus unexposed to childhood adversity (23), underscoring the importance of alternative exposure classifications.

In this study, we aim to extend these findings with methodological contributions that outline the structured life course modeling framework and its application in omics settings. As outlined in Dunn et al. (23), application of the SLCMA to omics data presents unique challenges not yet systematically investigated. First, it remains unknown whether theoretical properties of statistical inference, such as Type I error (i.e., family-wise error rate (FWER) in the presence of multiple testing) or confidence interval (CI) coverage, are valid in omics data. Second, it is unclear whether the SLCMA is sufficiently powered to detect the small effects commonly found in omics settings. Third, questions exist on how to balance decision-making regarding research evidence, because omics studies often rely on p-values and accurate statistical inference has become increasingly important. Moreover, epidemiologists and others increasingly prioritize other statistical evidence, such as effect sizes and CIs (24,25). We therefore performed simulations and empirical analyses to assess the performance of the SLCMA when applied to omics data, and illustrate how SLCMA can be applied to evaluate the time-dependent role of childhood abuse on genome-wide DNA methylation.

METHODS

The structured life course modeling approach (SLCMA): an overview

The SLCMA has been described in detail elsewhere (7,9,10). In brief, the SLCMA is a two-stage method that compares a set of life course hypotheses describing the relationship between exposures assessed over time and some outcome of interest. In the first stage of the SLCMA, each life course hypothesis is encoded into a predictor or set of predictor variables. **Table 1** shows examples of predictors representing commonly studied life course hypotheses. A variable selection procedure is then used to select the subset of predictors that explain the greatest proportion of outcome variation. While it is possible for multiple predictors to be selected, the high dimensionality of the omics setting makes consideration more feasible of simple life course hypotheses (meaning those in which the exposure-outcome association is represented by a single predictor). Therefore, in this study, we focused on statistical inference regarding the single predictor explaining the greatest variation in the outcome.

In the second stage of the SLCMA, post-selection inference is performed to obtain point estimates and CIs for the model identified from the first stage. Post-selection inference methods are used to derive unbiased test statistics because they account for the multiple testing that occurs when comparing multiple hypotheses (meaning, the multiple testing occurring at the first stage, instead of the number of outcomes examined), as the SLCMA iteratively works to *select* the variable with the strongest association with the outcome. Four inference methods that account for this "selective nature" are: Bonferroni correction; max-$|t|$-test (26); covariance test (27,28); and selective inference (29,30). These approaches are described in detail in **Web Appendix 1**.

Simulation analyses

We performed simulations to examine the performance of these four post-selection inference methods compared to a naïve calculation (summarized in **Table 2**). To build these simulations in the context real-world applications, we modeled the simulation strategy based on the genome-wide SLCMA study performed by Dunn et al. (23). We evaluated each post-selection inference method with respect to three statistical properties: family-wise error rate (FWER) (the probability of making one or more false discoveries out of multiple tests), statistical power (the probability of correctly selecting the predictor with a true association with the outcome), and CI coverage (the probability that a 95% CI contained the true effect estimate). Assessing these properties enabled us to determine whether inference based on these tests was valid in the presence of multiple testing and small effect sizes, which are two hallmarks of high-dimensional data. Mathematical definitions of the test-statistics and procedure for constructing CIs, as well as example R code are included in **Web Appendices 1 and 2.** All post-selection inference methods, including the naïve calculations, involved multiple testing correction for the number of cytosine-phosphate-guanine (CpG) sites tested using a Bonferroni correction (i.e., the p-value threshold was $p<1x10^{-7}$).

Simulations setup. We considered two scenarios, which differed in terms of the simulated outcome. In both scenarios, we simulated exposure to childhood sexual or physical abuse based on empirical data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a population-based birth cohort (31–33). The sample size was set to 700 to be consistent with ALSPAC. Simulations were based on m=485,000 tests corresponding to an analysis of Illumina Methylation 450k Beadchip data. In scenario 1, the outcome (i.e., DNA methylation) was simulated from a normal distribution. In scenario 2, we resampled the outcomes under the null to

more closely resemble 'beta' values, which represent the proportion of cells in which the cytosine at the locus is methylated and ranges from 0 to 1. To assess statistical power and CI coverage, we simulated the outcome from a beta distribution, as proposed by Tsai and Bell (31). In both scenarios, the effect sizes were selected to illustrate a wide range of statistical power based on previous epigenome-wide association studies examining different exposures (32,33).

To assess model misspecification, we also ran simulations in which: (1) the outcome variable was correlated with a variable encoding an alternative hypothesis (ever versus never exposed) not included in the prespecified set of hypotheses tested, (2) the outcome variable was correlated with two predictors (a compound life course hypothesis). We also varied the sample size to investigate its effect on statistical power.

Full details of simulations are provided in **Web Appendix 1**.

Measurement of power and confidence interval coverage. Conceptually, bias might arise from the SLCMA analysis in two ways: at the first stage, the model most supported by the sample data may not be the model most supported in the population. At the second stage, even if the model has been correctly selected, inference on that model may be biased. In our simulations, we considered both uncertainties residing in model selection and inference: power was calculated as the percentage of times that the first (variable selection) stage correctly selected the model and the second (inference) stage identified it as a below-threshold hit. Similarly, CI coverage was calculated as the percentage of times that the first stage correctly selected the model and the CI contains the true value. Alternatively, if the first stage selected the wrong model but the CI contains zero, we considered that the true effect (since there should be no effect) was captured by the CI.

Empirical analyses

To illustrate how the SLCMA and the different corresponding post-selection inference methods work in practice, we reanalyzed data used by Dunn et al. (23). Briefly, we compared the effects of sensitive period, accumulation, and recency hypotheses for the associations between exposure to sexual or physical abuse and genome-wide DNA methylation at age 7 in ALSPAC participants. Sample characteristics and adversity measures are described in **Web Appendix 3**. Building from that study, which only used the covariance test, we additionally applied the other post-selection inference methods summarized earlier.

The most widely used covariate adjustment strategy in the SLCMA is to regress the exposures on the covariates and enter the residuals into variable selection, which decreases the likelihood that observed associations are due to measured confounders. We also tested a new method for covariate adjustment that could be used alongside any post-selection inference method. Based on the Frisch-Waugh-Lovell theorem, this method also regresses the outcome on covariates and enters the residuals into the model selection procedure (34–36). A thorough description of this method and full list of covariates are available in **Web Appendix 1**. Of note, the SLCMA requires a common set of confounders to be pre-specified for all hypotheses; thus, bias may arise from time-varying or hypothesis-dependent confounding.

RESULTS

Simulation analyses

Table 3 summarizes the main findings from the simulation analyses regarding the statistical properties and implementation of the assessed methods.

Family-wise error rate. Due to the high computational burden of genome-wide association studies, we illustrated FWER control of each inference test using a single simulation with m=485,000 tests. As shown in **Figures 1 and 2**, when compared against the expected p-value distribution under the null hypothesis, the p-values obtained from naïve calculations appeared too liberal in both scenarios, as suggested by the systematic upward departure from the diagonal line. P-values from the covariance test were also smaller than expected across scenarios.

With normally distributed outcomes in Scenario 1, the p-values from the Bonferroni correction, max-$|t|$-test, and selective inference method followed the expected distribution closely (**Figure 1**). With empirical DNAm outcomes in Scenario 2, p-values from the three methods seemed conservative (**Figure 2**). Transforming the DNAm (beta) values to M-values did not affect the results (**Web Figure 1**). Together, these findings suggest that three methods adequately controlled the FWER: Bonferroni correction, the max-$|t|$-test, and the selective inference method. Estimates of FWER from repeated simulation experiments when the number of tests ranged from m=1 to 1,000 are available in **Web Appendix 1.**

Statistical power and CI coverage. We assessed the statistical power of the three methods that adequately controlled FWER. We did not evaluate the performance of the covariance test or

naïve calculation, as these methods would have their statistical power unfairly inflated by their tendency to fail to reject the null hypothesis.

Results suggested there was very little difference in statistical power between the three methods (**Figure 3**); they all had ideal statistical power (over 80%) when the effects were moderate to large ($R^2 > 0.06$ in scenario 1; $\Delta_{DNAm} > 0.25$ in scenario 2). With normal outcomes, the selective inference achieved ideal CI coverage (around 95%) across all effect sizes with sample size n = 700; the max-|*t*|-test had slightly lower coverage when the effect size was small ($R^2 < 0.03$). With beta distributed outcomes, the CI coverage probabilities were below the desired level (95%) when the between-group difference (Δ_{DNAm}) was below 0.3, though exceeded 95% as the effect size increased. Bonferroni corrected CIs were over-conservative across effect sizes and scenarios, as expected (**Figure 4**).

Robustness to model misspecification. If none of the predictors represent the true underlying life course hypothesis, then a misspecified model may be selected. In our simulations of this case, we found the accumulation or recency model were often selected, because they were highly correlated with the true predictor – ever versus never exposed ($r_{accumulation}$=0.89, $r_{recency}$=0.82). However, the power was reduced compared to a correctly specified model (**Web Figure 2)**. If the true hypothesis is represented by two or more predictors (*i.e.*, a compound hypothesis), then the power to select one of these predictors may be diminished. In our simulations, the power to select one predictor was lower for selective inference (**Figure 5**). However, selective inference is the only method available for post-selective inference on the second predictor that does not inflate FWER. Statistical power increased with sample size for all methods considered (**Web Figure 3**).

Empirical analyses

Using the covariance test, Dunn and colleagues identified five CpG sites in ALSPAC that showed differential methylation profiles at age 7 following exposure to physical or sexual abuse in childhood; the "sensitive period" model was the selected life course theory for these five sites. We reanalyzed the genome-wide SLCMA analyses using two other post-selection inference methods that showed no inflation in FWER and desired CI coverage: the max-|*t*|-test and selective inference method. Results are shown in **Web Table 1**. While neither method identified any CpG site as significantly associated using a stringent Bonferroni corrected p-value threshold of $p<1x10^{-7}$, the CpG site with the smallest p-value from the covariance test (*cg06430102*) remained the CpG with the smallest p-value (out of the 485,000 CpG sites tested) for the two alternative methods (**Web Table 1**). The CI calculated based on the covariance test, selective inference, and the max-|*t*|-test substantially overlapped (**Figure 6; Web Table 1**). On a genome-wide level, the concordance between the liberal covariance test and recommended selective inference method was high, implying that both methods agreed on the loci with the strongest associations with the exposure (**Web Table 2**).

After applying the Frisch-Waugh-Lovell theorem to additionally adjust for covariates, the p-values decreased at all five loci (**Web Figure 4**), suggesting that the approach improved statistical power while retaining control for confounding.

DISCUSSION

As the availability of longitudinal biological and phenotypic data grows in the era of big data, combining omics technologies with rigorous epidemiologic methods can reveal critical insights about biological mechanisms (37–39). Specifically, methods from life course epidemiology can be translated to "harness the 'omics' revolution" (2) and give insights to how exposures become biologically embedded. We showed that, under a set of untestable assumptions, one such method – the SLCMA – can be used to directly compare life course theories and scaled-up to answer nuanced questions about time-dependent exposure-omics relationships. For example, if an early childhood sensitive period hypothesis was selected for a locus known to be implicated in circadian rhythms, this finding could point to ways in which the biological clock is influenced by exposures during periods of heightened plasticity. If the accumulation hypothesis was selected for most of the loci implicated in inflammation, this finding could suggest dose-response relationships between the exposure and inflammatory responses.

Importantly, not all SLCMA methods for statistical inference are suitable in high-throughput applications. Our findings recommend the selective inference method and max-$|t|$-test for post-selection inference in omics applications. Our simulations also showed that statistical power to detect effects depended on effect size, but not necessarily on the post-selection inference method used. When deciding between these two inference methods, researchers will need to consider several factors, including goals of analysis and study-specific contexts, as both methods have strengths and limitations in these areas (**Web Appendix 1**). The simulation analyses highlight the value of using simulations in scientific research (40,41), especially when theoretical assumptions may be violated in a new application setting.

The empirical example in the current paper extended the analyses performed by Dunn et al. (23), using one of the exposures and the same DNAm data (**Web Appendix 3**). However, these analyses differed by considering two alternative post-selection inference methods (selective inference and the max-|*t*|-test) in the simulations. Comparing the covariance test to these two methods, we showed that statistical significance based on p-values may differ across methods. The main reason for the discordance between the max-|*t*|-test and the two lasso-based tests is that the max-|*t*|-test only considers the first predictor selected, whereas the selected inference is based on lasso models that also consider subsequent predictors. Researchers should assess p-values in parallel with effect estimates and CIs, as decision rules of significance based on p-values of one method may be biased due to inflation or overcorrection. Triangulating evidence from multiple sources and methods may suggest directions for future replication (42). For example, a CpG that was identified as the top site by multiple methods and showed substantial changes in methylation levels between exposed versus unexposed individuals may be more likely to capture effects of the exposure and worth pursuing in experimental validation.

Like any statistical method aspiring to address causal questions, the SLCMA relies on the usual assumptions that the model is correctly specified and that there is no unmeasured confounding (43). In simulations, we showed that when the model is misspecified, the SLCMA will identify hypothesized models with predictors that are correlated with the true model's predictors, but with reduced power. Therefore, SLCMA users must recognize that the selected hypothesis simply explains the most variation out of the (combinations of) candidate hypotheses considered, and there may be another (or non-tested) theoretical model that explains more variation. Thus, careful formulation of the hypotheses is critical to capture the most plausible causal relationship based on prior literature or reasoning; consideration of alternative hypotheses

(beyond those already selected) is also needed as research evidence grows. We would also emphasize that the selection of life course models is based both on proper specification of the relevant hypothesis and the set of candidate hypotheses included. For example, in our set of candidate hypotheses, we considered one sensitive period per time point when the exposure was measured. This approach may be inappropriate when the measurements are assessed close together in time: for example, for some exposure-outcome pairs we might not claim to distinguish a sensitive period at 1.5 years from one at 2.5 years. In such cases, we recommend condensing measurements into longer sensitive periods, taking the average exposure over all measurements in time period defined by prior literature or reasoning. Such an approach increases the statistical power of variable selection procedures by reducing the number of, and correlation between, predictors.

The SLCMA has some limitations beyond the usual assumptions: in the current study, we assumed the true hypothesis is represented by a single predictor (i.e., simple hypothesis). Identifying more complicated exposure-outcome relationships in the omics settings may be of interest but will require large sample sizes to achieve sufficient power. Moreover, the SLCMA currently does not accommodate time-varying confounding. It also does not allow for a different set of confounders for each hypothesis. In the empirical analyses, we tried to include a comprehensive set of baseline covariates based on prior literature that may be related to both childhood abuse and epigenetic changes. In light of these issues, the current results should be interpreted as suggestive evidence of loci that warrant future examination and replication in other datasets. Efforts to incorporate time-varying confounding into the SLCMA, such as marginal structural models (44,45), are also needed.

Several other limitations of the current study are noted. First, although we varied the effect size and compared normal versus empirical distributions of the outcome, we did not vary the distribution or correlation of the exposures, due to the number of possible combinations of these parameters. Thus, we encourage researchers to perform their own simulations to better understand the statistical properties of the SLCMA in their specific research context. Second, we restricted our analyses to linear regression-based model selection; a brief discussion on the possibility of implementing post-selection inference methods for generalized linear models is included in **Web Appendix 1**. Third, as suggested by the simulations, a typical longitudinal epigenetic study with a sample size under 1,000 is likely underpowered to detect small effects. In particular, when studying psychosocial exposures such as childhood abuse, we would not expect the exposure to have a large effect on DNAm at a single locus. For instance, power would likely be under 50% and CI coverage may be lower than 95% when the outcome variation explained is below 5%, which has been common in previous epigenome-wide association studies. One approach to improve statistical power is to combine data or summary results from multiple samples and perform a mega/meta-analysis; developing methods to meta-analyze results from SLCMA analyses is an important goal of future work. Another approach is to use the Frisch-Waugh-Lovell theorem for covariate adjustment, which, as shown in this paper, led to improvement in power. Fourth, the current SLCMA framework in the omics setting does not restrict or penalize any loci based on their biological significance. One promising direction of future research is to leverage functional or regulatory information about the genomic regions under consideration (46,47), especially when developmental stage-specific knowledge is available, in order to improve power and gain biological insights.

In conclusion, the SLCMA is a useful approach that brings the life course perspective into the omics context. Compared to an analysis that only categorized exposure status as exposed versus unexposed, the SLCMA not only offers additional mechanistic insights about exposure mechanisms, but also increases statistical power when the true underlying exposure-outcome relationship is more nuanced (23). As a field, we should move beyond analyses of the presence versus absence of exposure, and make full use of repeatedly measured phenotype and omics data to generate knowledge that improves human health over the life course.

Word count: 3,808 words

Acknowledgements:

We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. ARIES was funded by the BBSRC (BBI025751/1 and BB/I025263/1). Supplementary funding to generate DNA methylation data which is included in ARIES has been obtained from the MRC, ESRC, NIH and other sources. ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/2 and MC_UU_12013/8). A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). This publication is the work of the authors, each of whom serve as guarantors for the contents of this paper.

A part of this manuscript has been presented as a poster at the 35th International Society for Traumatic Stress Studies Annual Meeting in Boston, MA, 2019. It has also been submitted and approved as a poster presentation at the Society for Epidemiologic Research's 2020 Annual Meeting, scheduled to be held December 15-18, 2020, in Boston, MA.

References

1. De Stavola BL, Nitsch D, dos Santos Silva I, et al. Statistical Issues in Life Course Epidemiology. *Am J Epidemiol*. 2006;163(1):84–96.

2. Ben-Shlomo Y, Cooper R, Kuh D. The last two decades of life course epidemiology, and its relevance for research on ageing. *Int J Epidemiol*. 2016;45(4):973–988.

3. Kuh, D., Cooper, R., Hardy, R., Richards, M., & Ben-Shlomo, Y, eds. *A Life Course Approach to Healthy Ageing.* Oxford, UK: Oxford University Press; 2013.

4. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol.* 2002;31(2):285–293.

5. Kuh D. A Life Course Approach to Chronic Disease Epidemiology. Oxford University Press; 1997 344 p.

6. Koenen KC, Rudenstine S, Susser E, et al., eds. *A Life Course Approach to Mental Disorders.* Oxford, UK: Oxford University Press; 2013.

7. Mishra G, Nitsch D, Black S, et al. A structured approach to modelling the effects of binary exposure variables over the life course. *Int J Epidemiol*. 2009;38(2):528–537.

8. Efron B, Hastie T, Johnstone I, et al. Least angle regression. *The Annals of statistics*. 2004;32(2):407–499.

9. Smith ADAC, Heron J, Mishra G, et al. Model Selection of the Effect of Binary Exposures over the Life Course. *Epidemiology*. 2015;26(5):719–726.

10. Smith AD, Hardy R, Heron J, et al. A structured approach to hypotheses involving continuous exposures over the life course. *Int J Epidemiol*. 2016;45(4):1271–1279.

11. Madathil S, Joseph L, Hardy R, et al. A Bayesian approach to investigate life course hypotheses involving continuous exposures. *Int J Epidemiol*. 2018;47(5):1623-1635.

12. Dunn EC, Soare TW, Raffeld MR, et al. What life course theoretical models best explain the relationship between exposure to childhood adversity and psychopathology symptoms: recency, accumulation, or sensitive periods? *Psychological Medicine*. 2018;1–11.

13. Cooper R, Mishra GD, Kuh D. Physical Activity Across Adulthood and Physical Performance in Midlife: Findings from a British Birth Cohort. *American Journal of Preventive Medicine*. 2011;41(4):376–384.

14. Evans J, Melotti R, Heron J, et al. The timing of maternal depressive symptoms and child cognitive development: a longitudinal study. *Journal of Child Psychology and Psychiatry*. 2012;53(6):632–640.

15. Wills AK, Black S, Cooper R, et al. Life course body mass index and risk of knee osteoarthritis at the age of 53 years: evidence from the 1946 British birth cohort study. *Annals of the Rheumatic Diseases*. 2012;71(5):655–660.

16. Bann D, Kuh D, Wills AK, et al. Physical Activity Across Adulthood in Relation to Fat and Lean Body Mass in Early Old Age: Findings From the Medical Research Council National Survey of Health and Development, 1946–2010. *Am J Epidemiol*. 2014;179(10):1197–1207.

17. Dunn EC, Crawford KM, Soare TW, et al. Exposure to childhood adversity and deficits in emotion recognition: results from a large, population-based sample. *Journal of Child Psychology and Psychiatry*. 2018;59(8):845–854.

18. Nicolau B, Madathil SA, Castonguay G, et al. Shared social mechanisms underlying the risk of nine cancers: A life course study. *International Journal of Cancer*. 2019;144(1):59–67.

19. Huang JY, Gavin AR, Richardson TS, et al. Accounting for Life-Course Exposures in Epigenetic Biomarker Association Studies: Early Life Socioeconomic Position, Candidate Gene DNA Methylation, and Adult Cardiometabolic Risk. *Am J Epidemiol*. 2016;184(7):520–531.

20. Hughes A, Smart M, Gorrie-Stone T, et al. Socioeconomic Position and DNA Methylation Age Acceleration Across the Life Course. *Am J Epidemiol*. 2018;187(11):2346–2354.

21. Everson TM, Marsit CJ. Integrating -Omics Approaches into Human Population-Based Studies of Prenatal and Early-Life Exposures. *Curr Envir Health Rpt*. 2018;5(3):328–337.

22. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5–22.

23. Dunn EC, Soare TW, Zhu Y, et al. Sensitive Periods for the Effect of Childhood Adversity on DNA Methylation: Results From a Prospective, Longitudinal Study. *Biological Psychiatry*. 2019;85(10):838–849.

24. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305.

25. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "p < 0.05." *The American Statistician*. 2019;73(sup1):1–19.

26. Buja A, Brown L. Discussion: "A significance test for the lasso." *Ann. Statist.* 2014;42(2):509–517.

27. Lockhart R, Taylor J, Tibshirani RJ, et al. A significance test for the lasso. *Ann Stat*. 2014;42(2):413–468.

28. Lockhart R, Taylor J, Tibshirani R, et al. covTest: computes covariance test for adaptive linear modelling, R package version 1.02. (https://cran.r-project.org/web/packages/covTest/covTest.pdf).

29. Tibshirani RJ, Taylor J, Lockhart R, et al. Exact Post-Selection Inference for Sequential Regression Procedures. *Journal of the American Statistical Association*. 2016;111(514):600–620.

30. Tibshirani R, Tibshirani R, Taylor J, et al. selectiveInference: tools for post-selection inference, R package version 1.2.0. (https://cran.r-project.org/web/packages/selectiveInference/selectiveInference.pdf).

31. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol*. 2015;44(4):1429–1441.

32. Richmond RC, Simpkin AJ, Woodward G, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Human Molecular Genetics*. 2015;24(8):2201–2217.

33. Sharp GC, Lawlor DA, Richmond RC, et al. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*. 2015;44(4):1288–1304.

34. Frisch R, Waugh FV. Partial Time Regressions as Compared with Individual Trends. *Econometrica*. 1933;1(4):387.

35. Lovell MC. Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis. *Journal of the American Statistical Association*. 1963;58(304):993.

36. Yamada H. The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Communications in Statistics - Theory and Methods*. 2017;46(21):10897–10902.

37. Khoury MJ. A Primer Series on -Omic Technologies for the Practice of Epidemiology. *Am J Epidemiol*. 2014;180(2):127–128.

38. Khoury MJ. Planning for the Future of Epidemiology in the Era of Big Data and Precision Medicine. *Am J Epidemiol*. 2015;182(12):977–979.

39. Kuller LH. Epidemiologists of the Future: Data Collectors or Scientists? *Am J Epidemiol*. 2019;188(5):890–895.

40. König IR. Presidential address: Six open questions to genetic epidemiologists. *Genetic Epidemiology*. 2019;43(3):242–249.

41. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;38(11):2074–2102.

42. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6):1866–1886.

43. Howe LD, Smith AD, Macdonald-Wallis C, et al. Relationship between mediation analysis and the structured life course approach. *Int J Epidemiol*. 2016;45(4):1280-1294.

44. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*. 2000;11(5):550–560.

45. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008;168(6):656–664.

46. Kim S, Schliekelman P. Prioritizing hypothesis tests for high throughput data. *Bioinformatics*. 2016;32(6):850–858.

47. Iotchkova V, Ritchie GRS, Geihs M, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet*. 2019;51(2):343–353.

Table 1. Commonly Tested Life Course Theories.

Hypothesis	Life course theory	Definition	Encoding[a]	Example[b]
Sensitive period	The developmental timing of exposure *X* has the strongest effect on the outcome at a specific time point due to heightened levels of plasticity or reprogramming.	Exposure at a particular time point *j* (X_j) is associated with the outcome	X_j	$abuse_{period1}$ (X_1)=exposed (1) vs. unexposed (0) at time period 1
Accumulation	Every additional time point of exposure affects the outcome in a dose-response manner, independent of the exposure timing.	The accumulated sum of the number of exposure occasions (*A*) is linearly associated with the outcome.	$A = X_1 + \ldots + X_m$	$abuse_{accumulation}$ (*A*) =count of the number of time periods exposed to abuse (range 0-6)
Recency	More proximal exposures, meaning those that happen closer in time to the measurement of the outcome, are more strongly linked to the outcome compared to distal exposures.	The *weighted* sum (*R*) of the number of exposure occasions is linear associated with the outcome such that the weight of each exposure is proportional to the age at the time of measurement.	$R = X_1T_1 + \ldots + X_mT_m$	$abuse_{recency}$ (*R*) = $abuse_{period1}$ exposed (1) vs. unexposed (0)*($age_{period1}$) + …+ $abuse_{period6}$ exposed (1) vs. unexposed (0) *($age_{period6}$)
Mobility	The change in exposure status between two time periods, rather than the absolute state at each individual time point, affects the outcome.	The unidirectional change (M_{jk}^{+} or M_{jk}^{-}) between two measurement occasions (from *j*th to *k*th) is associated with the outcome.	Positive change: $M_{jk}^{+} = (1 - X_j)X_k$ Negative change: $M_{jk}^{-} = X_j(1 - X_k)$	$Abuse_{mobility^{+},period1to2}$ ($M_{1,2}^{+}$) = [1-exposed (1) at time period 1]*exposed(1) at time period 2

[a] The notations are based on the description of hypotheses by Smith et al.[9] Let $X_1, \ldots, X_m$ be a set of *m* repeated binary measures of exposure (0=unexposed; 1=exposed) and $T_1 \ldots T_m$ the corresponding age at the time of measurement. X_j represents the measure at the *j*th measurement occasion.

[b] Examples of how the life course theories could be encoded are shown in the last column, which were tested in the empirical analyses of epigenome-wide structured life course modeling approach (SLCMA) of exposure to physical or sexual abuse in childhood. Of note, the accumulation models can also be parameterized differently, such as with non-linear effects ("u-shaped" or "j-shaped" relationships). However, for simplicity, we provide the simplest definition of accumulation here, which is also often the most often tested.

Table 2. Summary of the Simulations Study Setup.

Under the null (Family-wise error rate)[a]			
	Predictors	**Outcome**	**Number of tests**
Normal outcomes	Based on exposure to childhood abuse from ALSPAC. Seven variables encoding sensitive period, accumulation and recency hypotheses.	$y \sim \mathcal{N}(0,1)$	485 000
Empirical outcomes	Based on exposure to childhood abuse from ALSPAC. Seven variables encoding sensitive period, accumulation and recency hypotheses.	Resampled DNAm values	485 000
Under the alternative (Power and confidence interval coverage)			
	Predictors	**Outcome**	**Effect size[b]**
Normal outcomes	Based on exposure to childhood abuse from ALSPAC. Seven variables encoding sensitive period, accumulation and recency hypotheses.	Simulated normal variables associated with the first predictor (earliest sensitive period)	R^2: 0.01 to 0.1
Empirical outcomes	Based on exposure to childhood abuse from ALSPAC. Seven variables encoding sensitive period, accumulation and recency hypotheses.	Simulated beta variables associated with the first predictor (earliest sensitive period)	Δ_{DNAm}: 0.05 to 0.5

[a] The table is divided into two approaches: to assess the family-wise error rate, we simulated the exposures and outcomes to have no association with each other (i.e., under the null hypothesis), and ran a single simulation of 485 000 tests to examine the distributions of observed p-values against the expected distribution. To assess the power and confidence interval coverage under the alternative hypothesis, we ran 2 000 simulation experiments to allow the confidence interval (CI) of the assessed metrics (i.e., power and CI coverage) to have a radius (i.e., margin of error) of 1%, setting α to 5%. The two metrics of effect sizes were different with normal versus empirical outcomes due to the difference in the underlying data generating processes. Sample size was set to n=700 in all simulations based on the sample size of the empirical study.

[b] R^2: variance of the outcome explained by the selected predictor; Δ_{DNAm}: difference in average methylation levels between the exposed and unexposed individuals

Table 3. Summary of Main Findings: Statistical Properties of Post-Selection Inference Methods.

Method	**FWER (Figures 1 & 2)**	**Statistical power (Figure 3)**	**CI coverage (Figure 4)**	**Software availability**	**Computation time for an epigenome-wide analysis[a]**
Naïve calculation	Inflated p-values and FWER	Biased due to inflated FWER	Lower than expected coverage when effect size is small[9]	Widely available	Fast (24 minutes)
Bonferroni correction	Controlled at any level	Comparable	Overconservative (i.e., above expected coverage)	Widely available	Fast (24 minutes)
Max-|*t*|-test	Controlled at any level	Comparable	Lower than expected coverage when effect size is small	R code provided in Web Appendix 2	Slow (11 hours 51 minutes)
Covariance test	Inflated p-values and FWER	Biased due to inflated FWER	Expected coverage;[9] interval not necessarily contiguous	R Package archived[28]	Moderate (1 hour 19 minutes)
Selective inference	Controlled at any level	Comparable	Expected coverage	R Package available;[30] possible to implement generalized linear models as well	Slow (14 hours 13 minutes)

FWER: family-wise error rate; CI: confidence interval

[a] Computation time was based on analyses running under R 3.4.0 using a high-performance computer cluster with 8GB RAM and a maximum of 6 CPU cores allotted.

Figure 1. Q-Q plots comparing the expected versus observed p-values simulated under the null for naïve calculations and four post-selection inference methods (N=700) with normal outcomes, where the outcome variables were simulated to follow a normal distribution (scenario 1). *Legend.*A) naïve calculations; B) covariance test (27); C) selective inference (29); D) max-|*t*|-test (26); E) Bonferroni correction

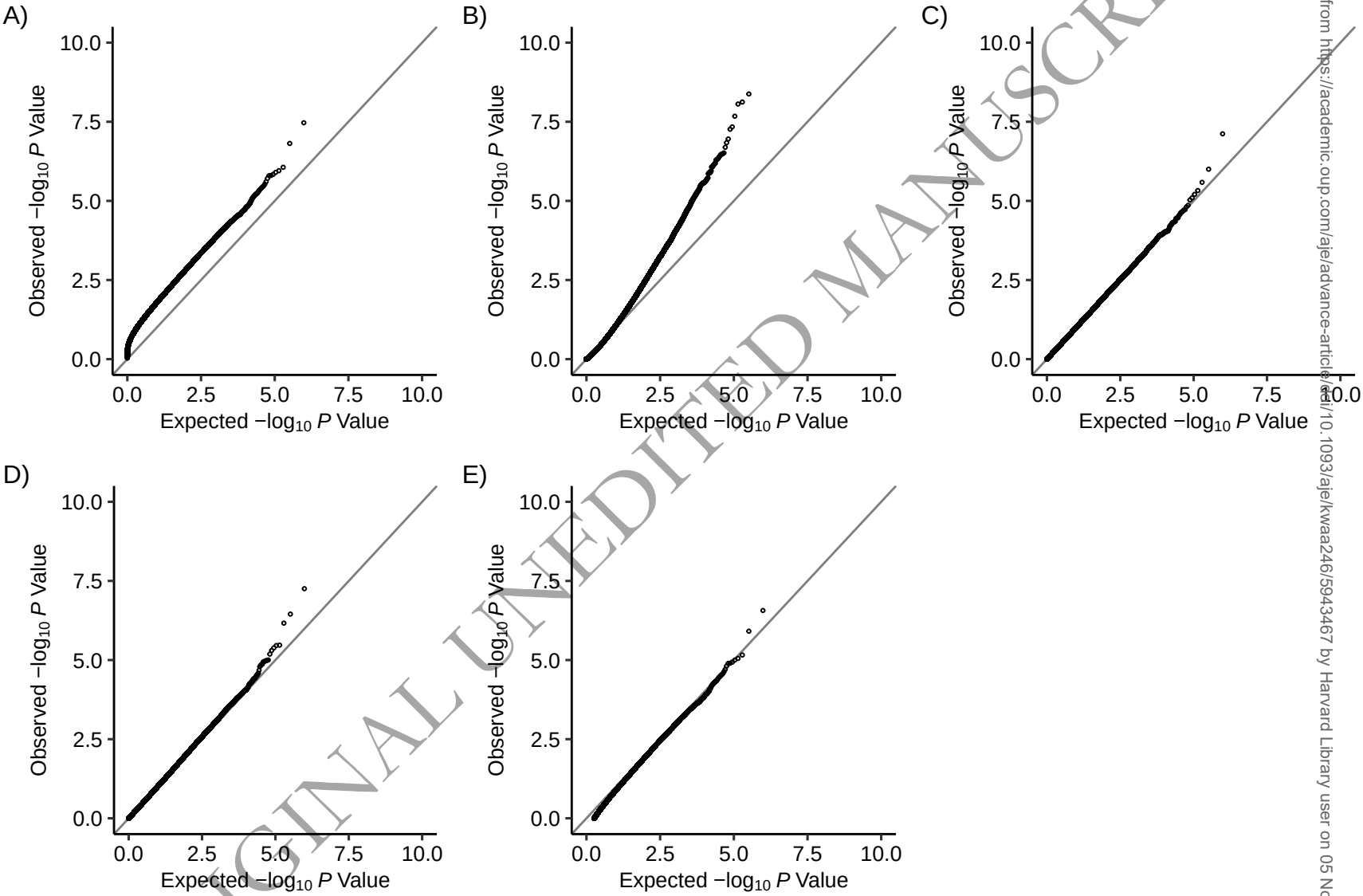

Figure 2. Q-Q plots comparing the expected versus observed p-values simulated under the null for naïve calculations and four post-selection inference methods (N=700) with empirical outcomes, where the outcome variables were resampled from observed DNAm values (scenario 2).

*Legend.*A) naïve calculations; B) covariance test (27); C) selective inference (29); D) max-|*t*|-test (26); E) Bonferroni correction

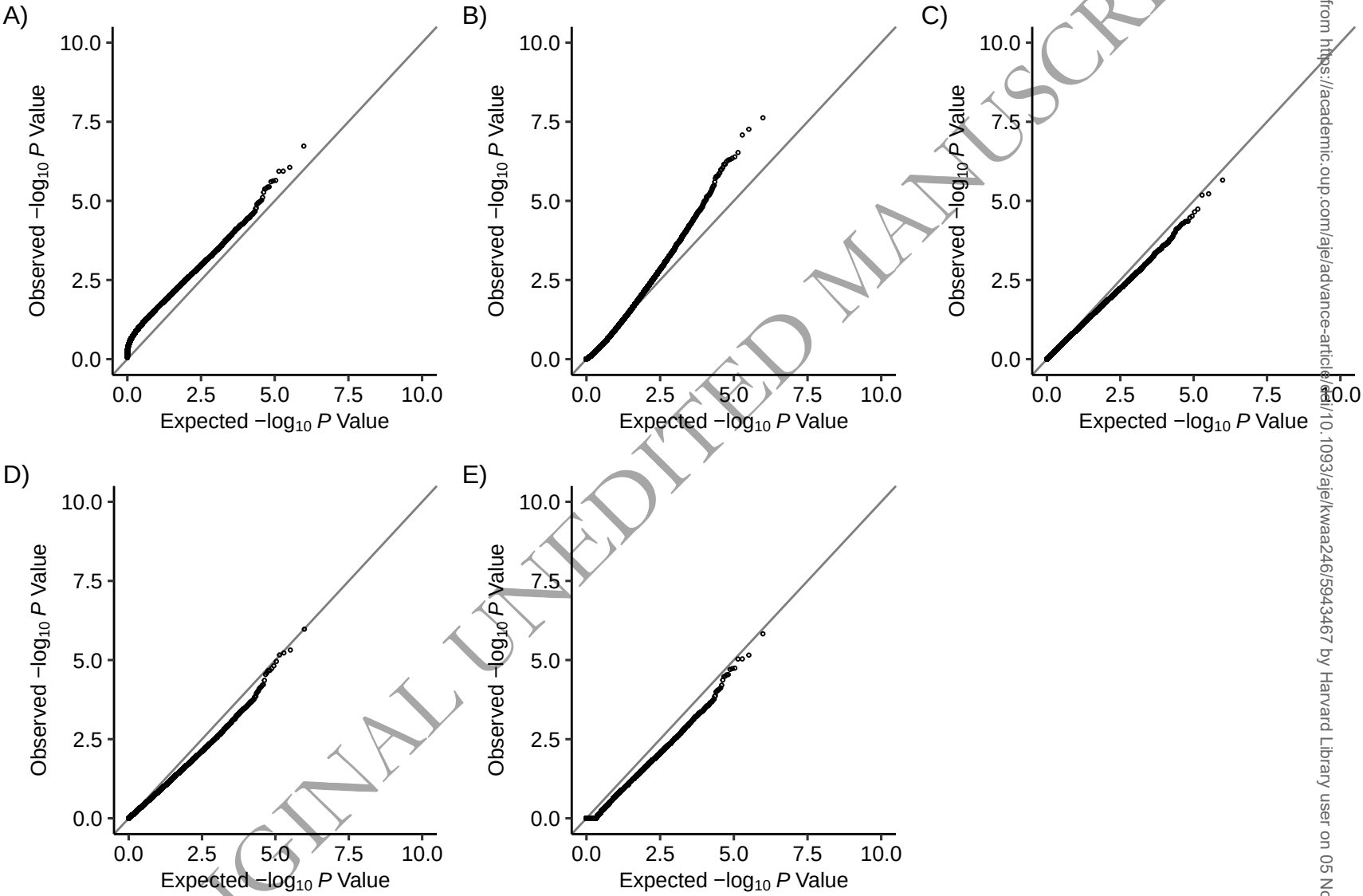

Figure 3. Estimated statistical power and corresponding 95% CI in simulated epigenome-wide analyses (n=700), with varying effect sizes.
Legend. A) normal outcomes; B) beta-distributed outcomes; Technical details about the selective inference (29) and max-|*t*|-test (26) are provided in Web Appendix 1.

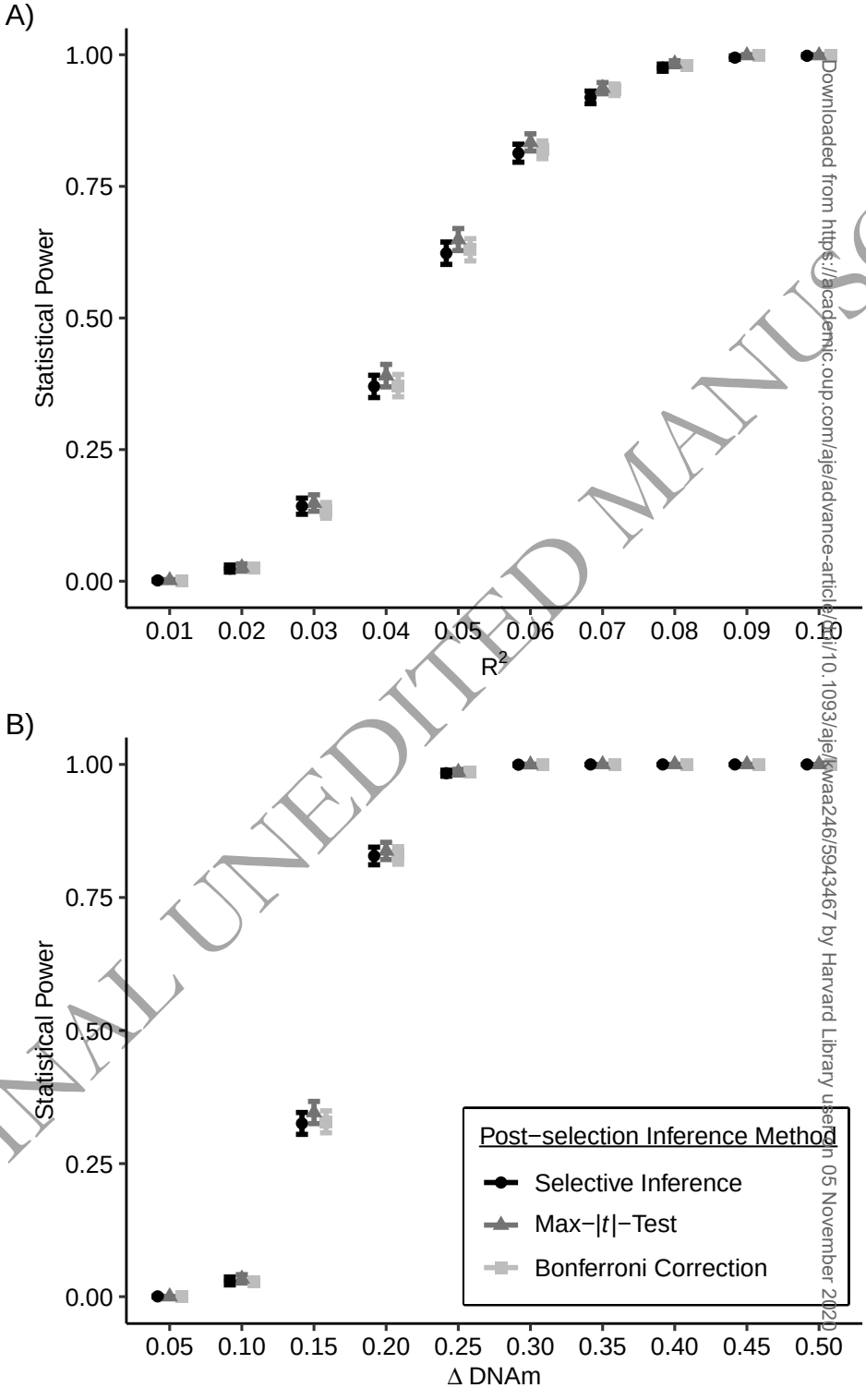

Figure 4. Estimated confidence interval coverage probability and corresponding 95% CI in simulated epigenome-wide analyses (n=700), with varying effect sizes. Gray dashed line corresponds to the pre-specified coverage probability (95%).
Legend. A) normal outcomes; B) beta-distributed outcomes; Technical details about the selective inference (29) and max-|*t*|-test (26) are provided in Web Appendix 1.

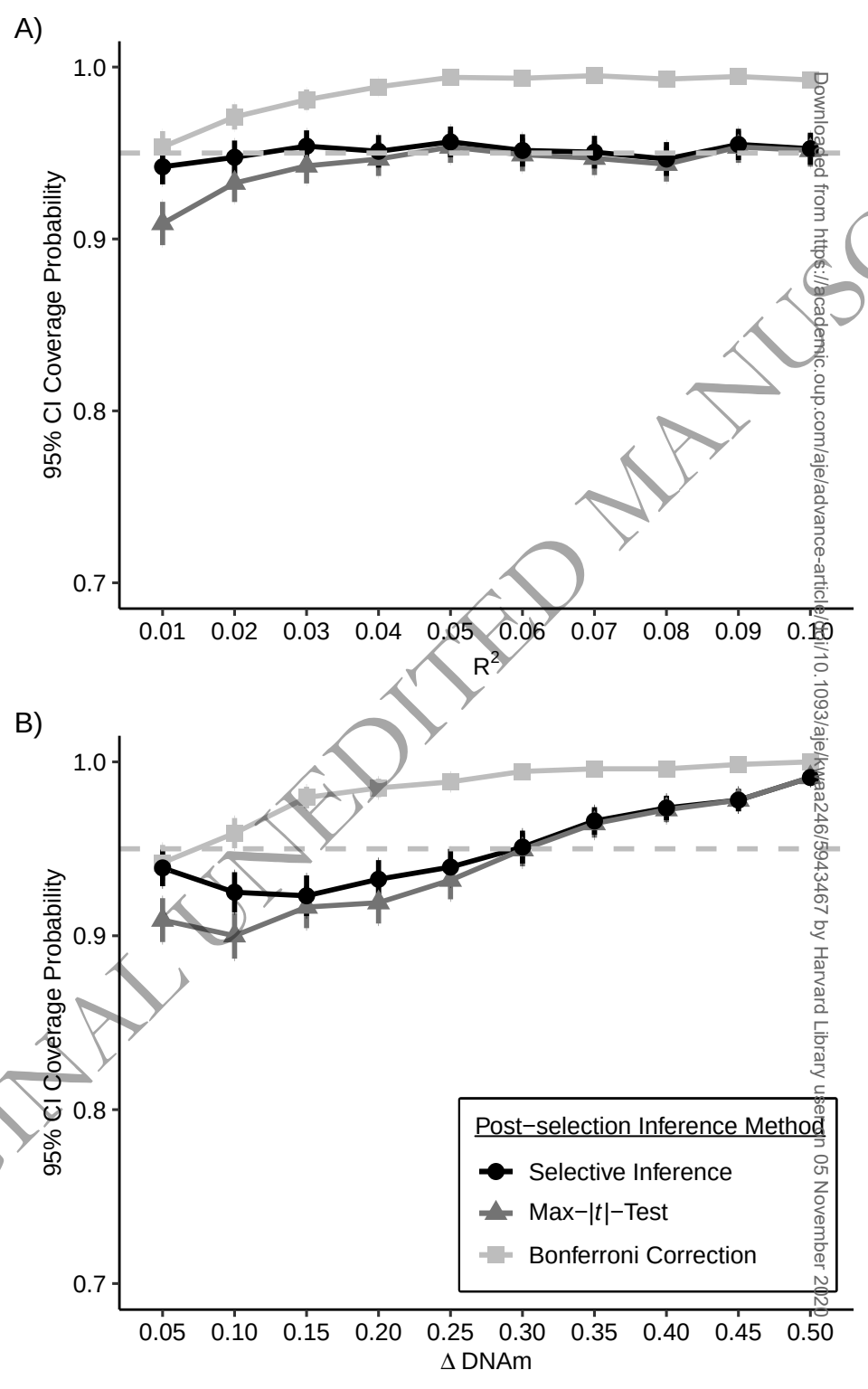

Figure 5. Estimated statistical power and corresponding 95% CI in simulated epigenome-wide analyses (n=700), with varying effect sizes, when the true causal relationship was represented by two hypotheses working in combination.
Legend. A) statistical power of selecting the first hypothesis (n=700), when the true hypothesis is a compound hypothesis; B) statistical power of selecting the second hypothesis (n=700), when the true hypothesis is a compound hypothesis; Technical details about the selective inference (29) and max-|*t*|-test (26) are provided in Web Appendix 1.

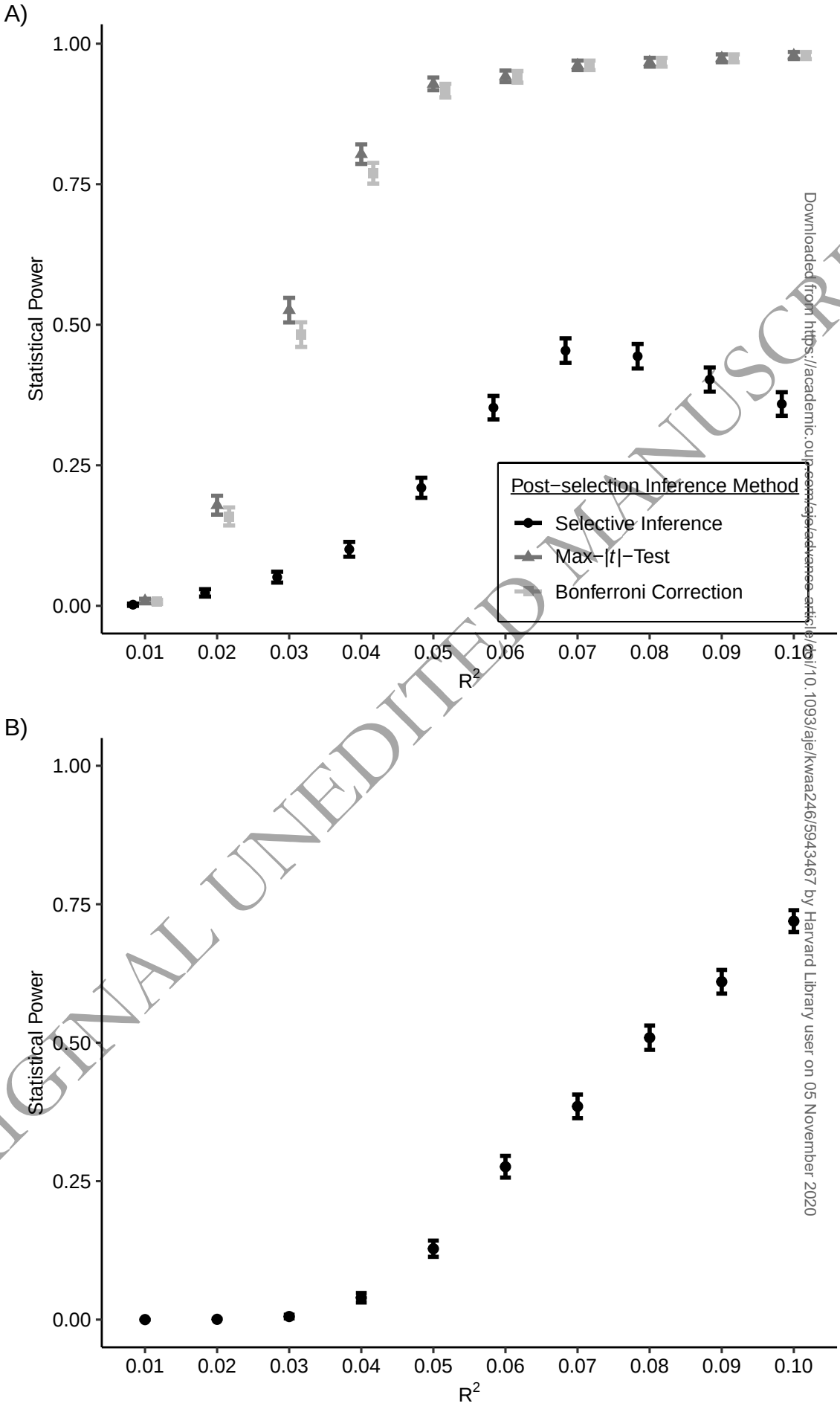

Figure 6. Overlap between confidence intervals based on the covariance test, selective inference, and the max-|t|-test in the empirical example, showing the top five loci.
Legend. Technical details about the covariance test (27), selective inference (29) and max-|*t*|-test (26) are provided in Web Appendix 1.