

1 A structured approach to evaluating life course hypotheses: Moving beyond analyses of exposed
2 versus unexposed in the omics context

3
4 Yiwen Zhu, MS^a

5 Andrew J. Simpkin, PhD^b

6 Matthew J. Suderman, PhD^c

7 Alexandre A. Lussier, PhD^{a,d}

8 Esther Walton, PhD^{c,e}

9 Erin C. Dunn, ScD, MPH^{*a,d,f,g,§}

10 Andrew D.A.C. Smith, PhD^{h,§}

11
12 ^aCenter for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

13 ^bSchool of Mathematics, Statistics and Applied Mathematics, National University of Ireland,
14 Galway, Ireland

15 ^cMRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of
16 Bristol, Bristol, UK

17 ^dDepartment of Psychiatry, Harvard Medical School, Boston, MA, USA

18 ^eDepartment of Psychology, University of Bath, Bath, UK

19 ^fStanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge,
20 MA, USA

21 ^gHenry and Alison McCance Center for Brain Health, Massachusetts General Hospital, Boston,
22 MA, USA

23 ^hApplied Statistics Group, University of the West of England, Bristol, UK

24
25 [§]Both senior authors contributed equally to this work.

26
27 **Corresponding Author:**

28 Erin C. Dunn, ScD, MPH, Psychiatric and Neurodevelopmental Genetics Unit, Center for
29 Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Simches Research
30 Building 6th Floor, Boston, MA 02114; email: [edunn\[at\]mgh\[dot\]harvard\[dot\]edu](mailto:edunn[at]mgh[dot]harvard[dot]edu). Phone: 617-
31 726-9387; Fax: 617-726-0830; website: www.thedunnlab.com

32
33 **Short running title:** Structured Life Course Modeling in Omics

34
35 **Funding:** This work was supported by the National Institute of Mental Health of the National
36 Institutes of Health [grant number R01MH113930 awarded to ECD]. The content is solely the
37 responsibility of the authors and does not necessarily represent the official views of the National
38 Institutes of Health.

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

53
54 **Conflict of Interest:** None declared.

55
56 **Word count: 2958 words**

57 **Figure number: 5**

58 **Table number: 3**

59

60

61 **Abstract**

62 **Background:** Life course epidemiology provides a framework for studying the effects of time-
63 varying exposures on health outcomes. The structured life course modeling approach (SLCMA)
64 is a theory-driven analytic method that empirically compares multiple prespecified life course
65 hypotheses characterizing time-dependent exposure-outcome relationships to determine which
66 theory best fits the observed data. However, the statistical properties of inference methods used
67 with the SLCMA have not been investigated with high-dimensional omics outcomes.

68
69 **Methods:** We performed simulations and empirical analyses to evaluate the performance of the
70 SLCMA when applied to genome-wide DNA methylation (DNAm). In the simulations, we
71 compared five statistical inference tests used by SLCMA (n=700). For each, we assessed the
72 familywise error rate (FWER), statistical power, and confidence interval coverage to determine
73 whether inference based on these tests was valid in the presence of substantial multiple testing
74 and small effect sizes, two hallmark challenges of inference from omics data. In the empirical
75 analyses, we applied the SLCMA to evaluate the time-dependent relationship of childhood abuse
76 with genome-wide DNAm (n=703).

77
78 **Results:** In the simulations, selective inference and max- $|t|$ -test performed best: both controlled
79 FWER and yielded moderate statistical power. Empirical analyses using SLCMA revealed time-
80 dependent effects of childhood abuse on DNA methylation.

81
82 **Conclusions:** Our findings show that SLCMA, applied and interpreted appropriately, can be
83 used in the omics setting to examine time-dependent effects underlying exposure-outcome

84 relationships over the life course. We provide recommendations for applying the SLCMA in
85 high-throughput settings, which we hope will encourage researchers to move beyond analyses of
86 exposed versus unexposed.

87

88 **Keywords:** Life course, structured approach, ALSPAC, omics, DNA methylation, post-
89 selection inference

90

91 **Key messages:**

- 92 1. The structured life course modeling approach (SLCMA) is an effective approach to
93 directly compare life course theories and can be scaled-up in the omics context to
94 examine nuanced relationships between environmental exposures over the life course and
95 biological processes.
- 96 2. Of the five statistical inference tests assessed in simulations, we recommend the selective
97 inference method and max- $|t|$ -test for post-selection inference in omics applications of the
98 SLCMA.
- 99 3. In an empirical example, we revealed time-dependent effects of childhood abuse on DNA
100 methylation using the SLCMA, with improvement in statistical power when accounting
101 for covariates by applying the Frisch-Waugh-Lovell (FWL) theorem.
- 102 4. Researchers should assess p-values in parallel with effects sizes and confidence intervals,
103 as triangulating multiple forms of statistical evidence can strengthen inferences and point
104 to new directions for replication.

105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127

Introduction

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 epidemiology¹ (**Table 1**), researchers are moving beyond simple comparisons of the presence versus absence of exposure to characterize time-dependent exposure-outcome relationships.² 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), the 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.⁷⁻¹¹ 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).³ 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 the empirical data. Smith and colleagues later extended this approach by outlining an alternative statistical model selection strategy that makes use of least angle regression (LARS),¹² accommodates both binary and continuous exposures,^{4,5} 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

128 theoretical hypothesis while assessing the relative importance of a series of life course theories
129 altogether.¹³ Since its inception, the SLCMA has been applied in a wide range of non-omic
130 epidemiologic studies, including those examining the time-dependent impacts of childhood
131 trauma, physical activity, or socioeconomic position on psychological, metabolic, and disease
132 outcomes.^{6,14-19}

133 The growing availability of high-dimensional biological and phenotype data from
134 longitudinal cohort studies has created new opportunities to test life course theories in
135 epigenomics, transcriptomics, metabolomics, and other omics settings. While large cross-
136 sectional omics studies have identified *associations* between biological differences and various
137 traits,²⁰ applications of the SLCMA to longitudinal data and high dimensional outcomes allow
138 researchers to answer more complex questions about disease *mechanisms*, including: Are there
139 sensitive periods when environmental exposures induce more underlying molecular changes
140 compared to those occurred during other time periods? Do exposures form a dose-response
141 relationship, regardless of their timing? Do more proximal events matter more compared to distal
142 events? As an example, Dunn and colleagues applied the SLCMA in a longitudinal birth cohort
143 study to model timing effects of childhood adversity on DNA methylation, which is a widely
144 studied epigenetic mechanism that could give rise to altered gene expression and phenotypic
145 changes. Using the SLCMA, they found that DNAm differences were largely explained by the
146 age at exposure, with the first three years of life appearing to be a sensitive period associated
147 with more DNAm differences. Interestingly, the SLCMA found associations not uncovered in
148 EWAS of exposed versus unexposed to childhood adversity.²²

149 As outlined in Dunn et al.,²² application of the SLCMA to omics data presents unique
150 challenges that have not been systematically investigated. First, it remains unknown whether

151 theoretical properties of statistical inference, such as Type I error (i.e., family-wise error rate
152 (FWER) in the presence of multiple testing) or confidence interval (CI) coverage, are valid in
153 omics data. Second, it is unclear whether the SLCMA is sufficiently powered to detect the small
154 effects commonly found in omics settings. Third, questions exist on how to balance decision-
155 making regarding research evidence; omics studies often rely on p-value thresholds whereas
156 epidemiologists and others increasingly prioritize other statistical evidence, such as effect sizes
157 and CIs.^{23,24} We therefore performed both simulations and empirical analyses to assess the
158 performance of the SLCMA when applied to omics data and illustrate how SLCMA can be
159 applied to evaluate the time-dependent role of childhood abuse on genome-wide DNA
160 methylation.

161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183

Methods

The Structured Life Course Modeling Approach (SLCMA): An Overview

The SLCMA has been described in detail elsewhere.³⁻⁵ 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. Examples of the predictors that represent commonly studied life course hypotheses are shown in **Table 1**. 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 that explains the greatest variation in the outcome.

In the second stage of the SLCMA, post-selection inference is performed to obtain point estimates (of the direction and magnitude of effects) and confidence intervals (a measure of effect uncertainty) of the effects for the hypotheses 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, as the SLCMA iteratively works to *select* the variable with the strongest association with the outcome. Four primary inference methods that account for this “selective nature” are: (1) Bonferroni correction; (2) max- $|t|$ -test;²⁵ (3) covariance test;^{26,27} and (4) selective inference.^{28,29} These approaches are described in detail in the **Supplementary Materials**.

184 **Simulation Analyses**

185 We performed simulations to examine the performance of these four post-selection
186 inference methods compared to a naïve calculation (summarized in **Table 2**). To build these
187 simulations in the context of a real-world application, we modeled the simulation strategy based
188 on the genome-wide SLCMA study performed by Dunn et al.²² We evaluated each post-selection
189 inference method with respect to three statistical properties: family-wise error rate (FWER) (the
190 probability of making one or more false discoveries out of multiple tests), statistical power (the
191 probability of correctly selecting the predictor with a true association with the outcome), and CI
192 coverage (the probability that a 95% CI contained the true effect estimate). Assessing these
193 properties enabled us to determine whether inference based on these tests was valid in the
194 presence of multiple testing and small effect sizes, which are two hallmarks of high-dimensional
195 data. Mathematical definitions of the test-statistics as well as the procedure for constructing
196 confidence intervals are included in the **Supplemental Materials**. Example R code for
197 implementing all methods described in this study are also provided in the **Supplemental**
198 **Materials**.

199 We considered two scenarios, which differed in the nature of the simulated outcome:

200

201 *Scenario 1: Normal outcomes*

202 In the first scenario, we simulated exposure to childhood sexual or physical abuse based
203 on empirical data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a
204 population-based birth cohort.^{30–32} The outcome was simulated from a normal distribution and
205 the sample size was set to 700 to be consistent with ALSPAC. Simulations were based on $m=485$
206 000 tests corresponding to an analysis of Illumina Methylation 450k Beadchip data and a

207 standard genome-wide significance threshold of $p < 10^{-7}$. To assess FWER under the null
208 hypothesis, we ran a single simulation of 485 000 tests and examined the distributions of
209 observed p-values against their expected distribution. To assess statistical power and CI
210 coverage, we ran simulations in which the outcome variable was correlated with one of the
211 predictors and then varied the correlation between outcome and predictor such that the variance
212 explained by the predictor varied from 0.01 to 0.1. This range of values was chosen to illustrate
213 poor and good statistical power, respectively.

214

215 *Scenario 2: Empirical outcomes*

216 To further examine the performance of post-selection inference methods in an empirical
217 setting, we combined real predictors from ALSPAC capturing repeatedly measured exposure to
218 childhood sexual or physical abuse as described above and an outcome variable that more closely
219 resembled the data distribution of DNAm at a single locus (meaning the proportion of cells in
220 which the cytosine at the locus is methylated). To assess the FWER under the null hypothesis,
221 we resampled the real predictors and DNAm values from ALSPAC separately. The resampling
222 breaks the predictor-outcome link and hence removes any observed association between the two,
223 while maintaining the empirical distributions of DNAm, which may not be normal. To assess
224 statistical power and CI coverage, we simulated the outcome from a beta distribution, as
225 proposed by Tsai and Bell.³³ Effect sizes were parameterized as the difference in mean levels of
226 DNAm between the exposed and unexposed groups ($\Delta DNAm$), ranging from 0.05 to 0.5. The
227 number of tests and p-value threshold were the same as Scenario 1. We additionally considered a
228 transformation of the DNAm values from beta values (y) to M values equivalent to
229 $M = \log_2(y/(1-y))$, which are sometimes used to stabilize variance.³⁴

230

231 **Empirical Analyses**

232 To illustrate how the SLCMA and the different corresponding post-selection inference
233 methods work in practice, we reanalyzed data used by Dunn et al.²² Briefly, we compared the
234 effects of sensitive period, accumulation, and recency hypotheses for the associations between
235 exposure to sexual or physical abuse and genome-wide DNA methylation at age 7 in ALSPAC
236 participants. Sample characteristics and adversity measures are described in **Supplemental**
237 **Materials**. Building from that study, which only used the covariance test, we additionally
238 applied the other three post-selection inference methods summarized earlier. We also tested a
239 new method for covariate adjustment that could be used alongside any post-selection inference
240 method. Based on the Frisch-Waugh-Lovell (FWL) theorem, this method regresses the outcome
241 on covariates in parallel and enters the residuals into the model selection procedure.^{35–37} A more
242 thorough description of this method and a full list of the covariates are available in the
243 **Supplemental Materials**.

244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266

Results

Simulation Analyses

Table 3 summarizes the main findings from the simulation analyses with regard to 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 illustrate 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 too conservative (**Figure 2**). Transforming the DNAm (beta) values to M -values did not affect the results (**Figure S1**). 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 1000 are available in the **Supplemental Materials**.

Statistical Power and CI Coverage

267 We assessed the statistical power of the three methods that adequately controlled FWER.
268 We did not evaluate the performance of the covariance test or naïve calculation, as these methods
269 had smaller than expected p-values under the null hypothesis in all simulations. Inclusion of
270 these methods would have thus resulted in an unfair comparison, as their statistical power is
271 inflated by their tendency to fail to reject the null hypothesis.

272 **Figure 3** shows the estimated statistical power as a function of effect sizes tested both
273 with normal and empirical outcomes. There was very little difference in statistical power
274 between the three methods; they all had ideal statistical power (over 80%) when the effects were
275 moderate to large ($R_2 > 0.06$ in scenario 1; $\Delta DNAm > 0.25$ in scenario 2).

276 **Figure 4** shows CI coverage as a function of effect size. With normal outcomes, the
277 selective inference achieved ideal coverage (around 95%) across all effect sizes with sample size
278 $n = 700$; the max- $|t|$ -test had slightly lower coverage when the effect size was small ($R_2 < 0.03$).
279 With empirical (beta distributed) outcomes, the CI coverage probabilities were below the desired
280 level (95%) when the between-group difference ($\Delta DNAm$) was below 0.3, though exceeded 95%
281 as the effect size increased. Bonferroni corrected CIs were over-conservative across effect sizes
282 and scenarios, as expected.

283

284 **Empirical Analyses**

285 Using the covariance test, Dunn and colleagues identified five CpG sites in ALSPAC that
286 showed differential methylation profiles at age 7 following exposure to physical or sexual abuse
287 in childhood; the “sensitive period” model was the selected life course theory for these five sites.
288 We reanalyzed the genome-wide SLCMA analyses using two other post-selection inference
289 methods that showed no inflation in FWER and desired CI coverage: the max- $|t|$ -test and

290 selective inference method. Results are shown in **Table S1**. While neither method identified any
291 CpG site as significantly associated using a stringent Bonferroni corrected p-value threshold of
292 $p < 1 \times 10^{-7}$, the CpG site with the smallest p-value from the covariance test (*cg06430102*)
293 remained the CpG with the smallest p-value (out of the 485 000 CpG sites tested) for the two
294 alternative methods (**Table S1**). The CI calculated based on the covariance test, selective
295 inference, and the max- $|t|$ -test substantially overlapped (**Figure 5; Table S1**).

296 After applying the Frisch-Waugh-Lovell (FWL) theorem to additionally adjust for
297 covariates, the p-values decreased at all five loci (**Figure S2**), suggesting that the approach
298 improved statistical power and additionally controlled for any bias that may be caused by
299 confounding.

300 On a genome-wide level, **Table S2** shows the overlap in most strongly associated loci
301 based on results obtained from the three methods assessed in the empirical analyses to compare
302 the preferred alternative methods to the widely-used (but now identified as inflated) covariance
303 test. The concordance between the liberal covariance test and recommended selective inference
304 method was high, implying that both methods agreed on the loci with the strongest associations
305 with the exposure.

306

Discussion

307 As the availability of longitudinal biological and phenotypic data grows, methods from
308 life course epidemiology can be translated to “harness the ‘omics’ revolution”⁷ and reveal
309 critical insights about how exposures become biologically embedded. We showed that one such
310 method – the SLCMA – is an effective approach to directly compare life course theories and can
311 be scaled-up to answer nuanced questions about complex and time-dependent exposure-omics
312 relationships. Importantly, not all methods for statistical inference in the SLCMA are suitable in
313 high-throughput applications. Based on our findings, we recommend the selective inference
314 method and max- $|t|$ -test for post-selection inference in omics applications. Our simulations also
315 showed that statistical power to detect effects depended on effect size, but not necessarily on the
316 post-selection inference method used. When deciding between these two inference methods,
317 researchers will need to consider several factors, including goals of analysis and study-specific
318 contexts, as both methods have strengths and limitations in these areas (**Supplemental**
319 **Materials**). The simulation analyses highlight the value of using simulations in scientific
320 research,^{39,40} especially when theoretical assumptions may be violated in a new application
321 setting.

322 In the empirical analyses, we also showed that evaluating multiple forms of evidence can
323 help contextualize the SLCMA findings in omics applications, though statistical significance
324 based on p-values may differ across methods. Researchers should assess p-values in parallel with
325 effects sizes and confidence intervals, as decision rules of significance based on p-values of one
326 method may be biased due to inflation or overcorrection. While it remains challenging to draw
327 definitive conclusions about the presence versus absence of an effect, triangulating evidence
328 from multiple sources and methods may suggest directions for future replication.⁴¹ For example,

329 a CpG that was identified as the top site by multiple methods and showed substantial changes in
330 methylation levels between exposed versus unexposed individuals may be more likely to capture
331 effects of the exposure and worth pursuing in experimental validation.

332 Several limitations are noted. First, although we varied the effect size and compared
333 normal versus empirical distributions of the outcome, we did not assess diverse correlation
334 structures or distributions in the exposures tested, as it was not possible given the number of
335 possible combinations of these parameters. Thus, we encourage researchers to perform their
336 own simulations to arrive at a better understanding of the statistical properties of the SLCMA in
337 their specific research context. Second, we restricted our analyses to the application of linear
338 regression-based model selection; a brief discussion on the possibility of implementing post-
339 selection inference methods for generalized linear models is included in the **Supplemental**
340 **Materials**. Third, as suggested by the simulations, a typical longitudinal epigenetic study with a
341 sample size under 1000 is likely underpowered to detect small effects. One approach to improve
342 statistical power is to combine data or summary results from multiple samples and perform a
343 mega/meta-analysis; developing methods to meta-analyze results from SLCMA analyses is an
344 important goal of future work. Another approach is to use the FWL theorem for covariate
345 adjustment, which as shown in this paper led to improvement in power.

346 In conclusion, the SLCMA is a useful approach that brings the life course perspective
347 into the omics context. This analysis framework can reveal site-specific mechanisms, as well as
348 system-level pathways. Compared to an analysis that only categorized exposure status as
349 exposed versus unexposed, the SLCMA not only offers additional mechanistic insights, but also
350 increases statistical power when the true underlying exposure-outcome relationship is more
351 nuanced.²² As a field, we should move beyond analyses of the presence versus absence of

352 exposure, and make full use of repeatedly measured phenotype and omics data to generate
353 knowledge that improves human health over the life course.

References

1. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health*. 2003 Oct;**57**(10):778–783.
2. De Stavola BL, Nitsch D, Santos Silva I dos, et al. Statistical Issues in Life Course Epidemiology. *Am J Epidemiol*. 2006 Jan 1;**163**(1):84–96.
3. Mishra G, Nitsch D, Black S, De Stavola B, Kuh D, Hardy R. A structured approach to modelling the effects of binary exposure variables over the life course. *Int J Epidemiol*. 2009 Apr 1;**38**(2):528–537.
4. Smith ADAC, Heron J, Mishra G, Gilthorpe MS, Ben-Shlomo Y, Tilling K. Model Selection of the Effect of Binary Exposures over the Life Course. *Epidemiology*. 2015 Sep;**26**(5):719–726.
5. Smith AD, Hardy R, Heron J, et al. A structured approach to hypotheses involving continuous exposures over the life course. *Int J Epidemiol*. 2016 Aug 1;**45**(4):1271–1279.
6. 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 Feb;1–11.
7. 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 Aug 1;**45**(4):973–988.
8. Ben-Shlomo Y. A Life Course Approach to Healthy Ageing [Internet]. Oxford University Press; 2013 [cited 2019 Jul 8]. Available from: <https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199656516.001.0001/acprof-9780199656516>
9. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 2002 Apr 1;**31**(2):285–293.
10. Kuh D. A Life Course Approach to Chronic Disease Epidemiology. Oxford University Press; 1997.
11. Koenen KC, Rudenstine S, Susser E, Galea S. A Life Course Approach to Mental Disorders. OUP Oxford; 2013.
12. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *The Annals of statistics*. 2004;**32**(2):407–499.
13. Madathil S, Joseph L, Hardy R, Rousseau M-C, Nicolau B. A Bayesian approach to investigate life course hypotheses involving continuous exposures. *Int J Epidemiol* [Internet]. [cited 2018 Jun 28]; Available from: <https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyy107/5038105>

14. 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 Oct 1;**41**(4):376–384.
15. 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.
16. 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 May 1;**71**(5):655–660.
17. Bann D, Kuh D, Wills AK, Adams J, Brage S, Cooper R. 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 May 15;**179**(10):1197–1207.
18. 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.
19. Nicolau B, Madathil SA, Castonguay G, Rousseau M-C, Parent M-E, Siemiatycki J. Shared social mechanisms underlying the risk of nine cancers: A life course study. *International Journal of Cancer*. 2019;**144**(1):59–67.
20. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017 Jul 6;**101**(1):5–22.
21. 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 Apr 15;**24**(8):2201–2217.
22. 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 May 15;**85**(10):838–849.
23. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019 Mar;**567**(7748):305.
24. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond “ $p < 0.05$ ”. *The American Statistician*. 2019 Mar 29;**73**(sup1):1–19.
25. Buja A, Brown L. Discussion: “A significance test for the lasso”. *Ann Statist*. 2014 Apr;**42**(2):509–517.
26. Lockhart R, Taylor J, Tibshirani RJ, Tibshirani R. A significance test for the lasso. *Ann Stat*. 2014 Apr;**42**(2):413–468.

27. Lockhart R, Taylor J, Tibshirani R, Tibshirani R. covTest: computes covariance test for adaptive linear modelling, R package version 1.02 [Internet]. [cited 2019 Jan 23]. Available from: <https://cran.r-project.org/web/packages/covTest/covTest.pdf>
28. Tibshirani RJ, Taylor J, Lockhart R, Tibshirani R. Exact Post-Selection Inference for Sequential Regression Procedures. *Journal of the American Statistical Association*. 2016 Apr 2;**111**(514):600–620.
29. Tibshirani R, Tibshirani R, Taylor J, Loftus J, Reid S. selectiveInference: tools for post-selection inference, R package version 1.2.0 [Internet]. [cited 2019 Jan 23]. Available from: <https://cran.r-project.org/web/packages/selectiveInference/selectiveInference.pdf>
30. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013 Feb;**42**(1):97–110.
31. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The ‘Children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013 Feb;**42**(1):111–127.
32. Relton CL, Gaunt T, McArdle W, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol*. 2015 Aug;**44**(4):1181–1190.
33. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol*. 2015 Aug;**44**(4):1429–1441.
34. Du P, Zhang X, Huang C-C, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* [Internet]. 2010 Dec [cited 2019 Apr 23];**11**(1). Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-587>
35. Frisch R, Waugh FV. Partial Time Regressions as Compared with Individual Trends. *Econometrica*. 1933 Oct;**1**(4):387.
36. Lovell MC. Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis. *Journal of the American Statistical Association*. 1963 Dec;**58**(304):993.
37. Yamada H. The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Communications in Statistics - Theory and Methods*. 2017 Nov 2;**46**(21):10897–10902.
38. Saffari A, Silver MJ, Zavattari P, et al. Estimation of a significance threshold for epigenome-wide association studies. *Genet Epidemiol*. 2018 Feb 1;**42**(1):20–33.
39. König IR. Presidential address: Six open questions to genetic epidemiologists. *Genetic Epidemiology*. 2019;**43**(3):242–249.
40. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;**38**(11):2074–2102.

41. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol.* 2016 01;**45**(6):1866–1886.

Table 1. Commonly tested life course theories.

Hypothesis	Life course theory	Definition	Encoding	Example
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	$\text{abuse}_{\text{period1}}(X_j) = \text{exposed (1) vs. unexposed (0) at time period 1}$
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.	$A = X_1 + \dots + X_m$	$\text{abuse}_{\text{accumulation}}(A) = \text{count of the number of time periods exposed to abuse (range 0-6)}$
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 <i>weighted</i> 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.	$R = X_1T_1 + \dots + X_mT_m$	$\text{abuse}_{\text{recency}}(R) = \text{abuse}_{\text{period1}} \text{ exposed (1) vs. unexposed (0)} * (\text{age}_{\text{period1}}) + \dots + \text{abuse}_{\text{period6}} \text{ exposed (1) vs. unexposed (0)} * (\text{age}_{\text{period6}})$
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: $M_{jk}^- = X_j(1 - X_k)$	$\text{Abuse}_{\text{mobility+}, \text{period1to2}}(M_{1,2}^+) = [1 - \text{exposed (1) at time period 1}] * \text{exposed(1) at time period 2}$

The notations are based on the description of hypotheses by Smith et al.⁴ Let X_1, \dots, X_m be a set of m repeated binary measures of exposure (0=unexposed; 1=exposed) and $T_1 \dots T_m$ the corresponding age at the time of measurement. X_j represents the measure at the j th measurement occasion. 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 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)			
	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		Resampled DNAm values	485 000
Under the alternative (Power and confidence interval coverage)			
	Predictors	Outcome	Effect size
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		Simulated beta variables associated with the first predictor (earliest sensitive period)	Δ_{DNAm} : 0.05 to 0.5

The table is divided into two approaches: to assess the familywise 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.

R₂: 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*
Naïve calculation	Inflated p-values and FWER	Biased due to inflated FWER	Lower than expected coverage when effect size is small ⁴	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 the Supplement	Slow (11 hours 51 minutes)
Covariance test	Inflated p-values and FWER	Biased due to inflated FWER	Expected coverage; ⁴ interval not necessarily contiguous	R Package archived ²⁷	Moderate (1 hour 19 minutes)
Selective inference	Controlled at any level	Comparable	Expected coverage	R Package available; ²⁹ possible to implement generalized linear models as well	Slow (14 hours 13 minutes)

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

*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).

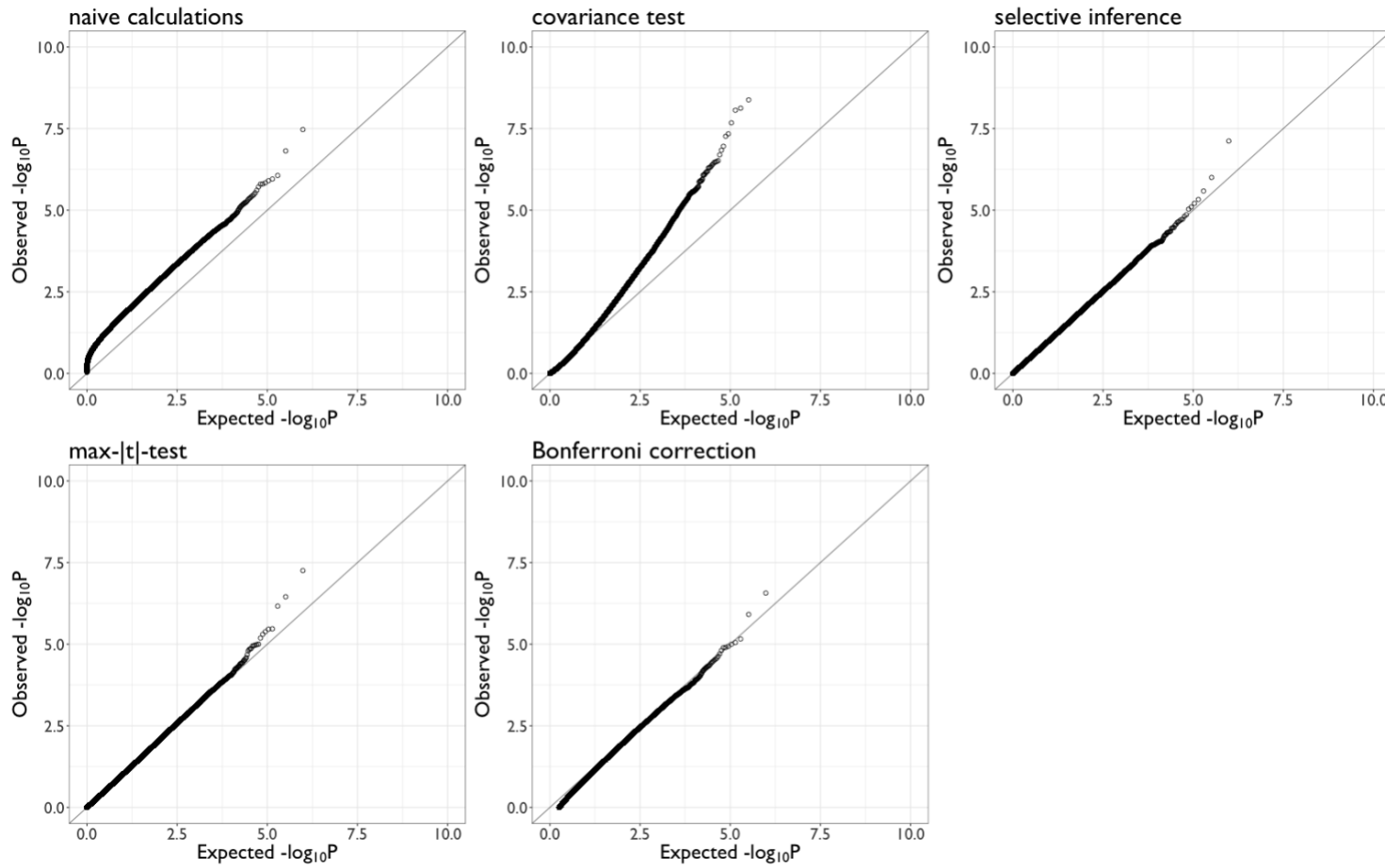


Figure 2. Q-Q plots comparing the expected versus observed p-values simulated under the null for naïve calculations and four post-selection inference methods (N=700) with empirical outcomes, where the outcome variables were resampled from observed DNAm values (scenario 2).

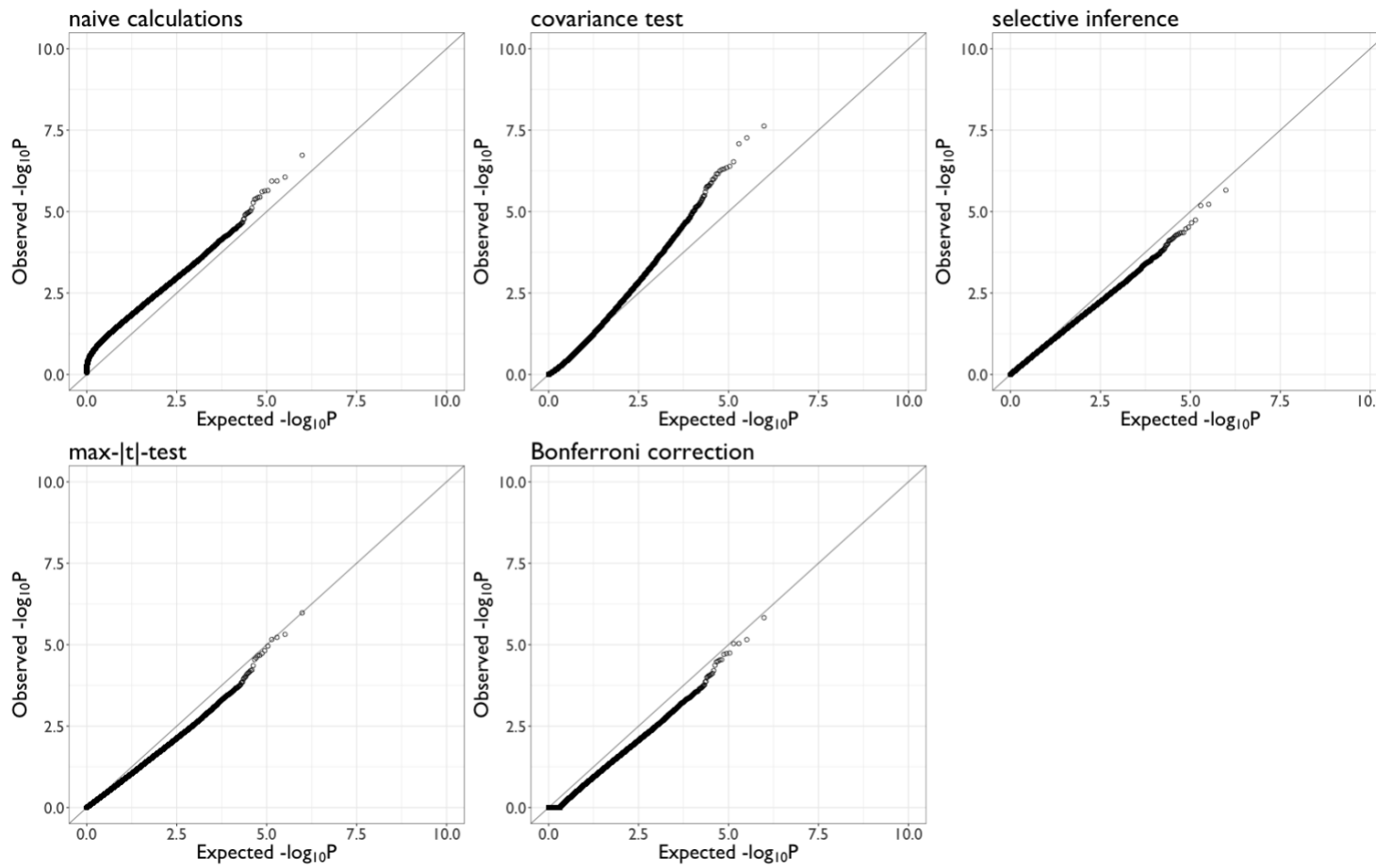


Figure 3. Estimated statistical power and corresponding 95% CI in simulated epigenome-wide analyses (n=700), with varying effect sizes.

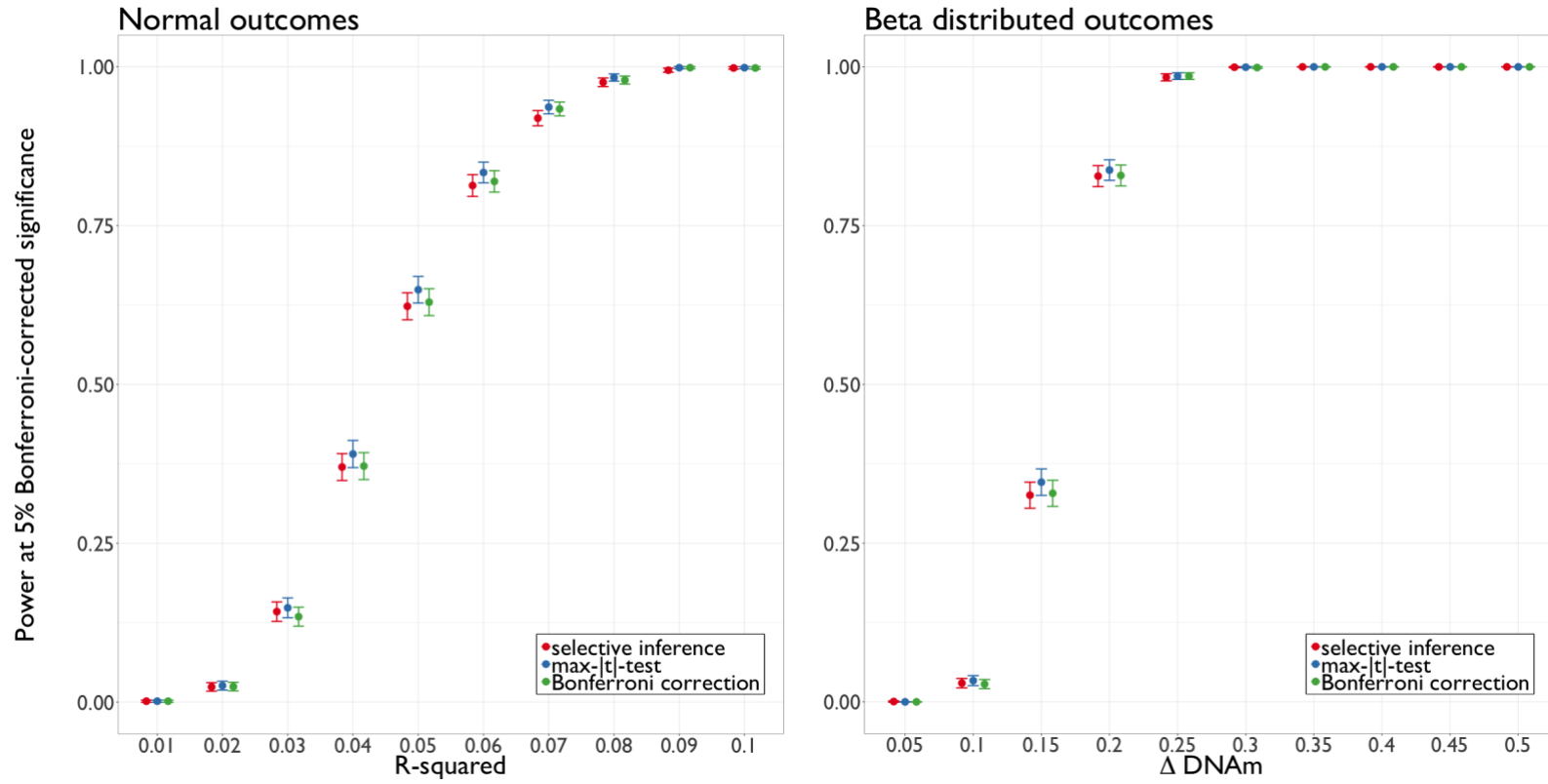
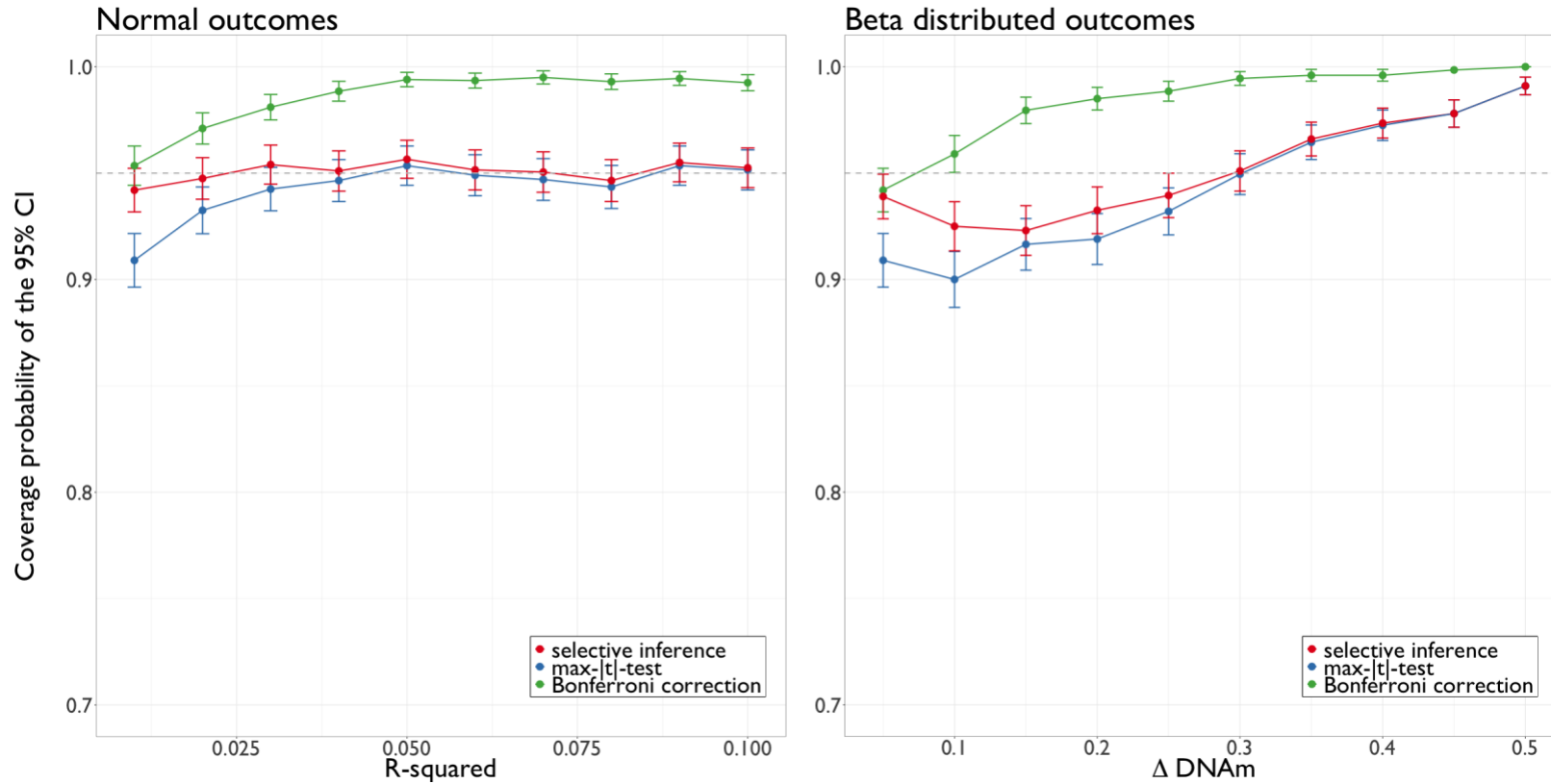


Figure 4. Estimated confidence interval coverage probability and corresponding 95% CI in simulated epigenome-wide analyses (n=700), with varying effect sizes.



Gray dashed line corresponds to the pre-specified coverage probability (95%).

Figure 5. 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.

