

University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2005

Paper 168

Multiple Testing Procedures and Applications to Genomics

Merrill D. Birkner*

Katherine S. Pollard[†]

Mark J. van der Laan[‡]

Sandrine Dudoit**

*Division of Biostatistics, School of Public Health, University of California, Berkeley, mbirkner@berkeley.edu

[†]Center for Molecular Science & Engineering, University of California, Santa Cruz, kpollard@gladstone.ucsf.edu

[‡]Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

**Division of Biostatistics, School of Public Health, University of California, Berkeley, sandrine@stat.berkeley.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/ucbbiostat/paper168>

Copyright ©2005 by the authors.

Multiple Testing Procedures and Applications to Genomics

Merrill D. Birkner, Katherine S. Pollard, Mark J. van der Laan, and Sandrine Dudoit

Abstract

This chapter proposes widely applicable resampling-based single-step and step-wise multiple testing procedures (MTP) for controlling a broad class of Type I error rates, in testing problems involving general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics (Dudoit and van der Laan, 2005; Dudoit et al., 2004a,b; van der Laan et al., 2004a,b; Pollard and van der Laan, 2004; Pollard et al., 2005). Procedures are provided to control Type I error rates defined as tail probabilities for arbitrary functions of the numbers of Type I errors, V_n , and rejected hypotheses, R_n . These error rates include: the generalized family-wise error rate, $gFWER(k) = \Pr(V_n > k)$, or chance of at least $(k+1)$ false positives (the special case $k=0$ corresponds to the usual family-wise error rate, FWER), and tail probabilities for the proportion of false positives among the rejected hypotheses, $TPFP(q) = \Pr(V_n/R_n > q)$. Single-step and step-down common-cut-off (maxT) and common-quantile (minP) procedures, that take into account the joint distribution of the test statistics, are proposed to control the FWER. In addition, augmentation multiple testing procedures are provided to control the gFWER and TPFP, based on any initial FWER-controlling procedure. The results of a multiple testing procedure can be summarized using rejection regions for the test statistics, confidence regions for the parameters of interest, or adjusted p-values. A key ingredient of our proposed MTPs is the test statistics null distribution (and consistent bootstrap estimator thereof) used to derive rejection regions and corresponding confidence regions and adjusted p-values. This chapter illustrates an implementation in SAS (Version 9) of the bootstrap-based single-step maxT procedure and of the gFWER- and TPFP-controlling augmentation procedures. These multiple testing procedures are applied to an HIV-1 sequence dataset to identify codon positions associated with viral replication capacity.

Contents

1	Introduction	2
1.1	Motivation	2
1.2	Outline	4
2	Multiple hypothesis testing methodology	4
2.1	Multiple hypothesis testing framework	4
2.2	Test statistics null distribution	10
2.2.1	Characterization of the test statistics null distribution .	10
2.2.2	Bootstrap estimation of the test statistics null distribution	12
2.3	Rejection regions	14
2.4	Single-step procedures for controlling general Type I error rates $\theta(F_{V_n})$	15
2.5	Step-down procedures for controlling the family-wise error rate	17
2.6	Augmentation multiple testing procedures for controlling tail probability error rates	18
3	Software implementation in SAS	20
4	Application to HIV-1 sequence data	22
4.1	HIV-1 dataset	23
4.1.1	HIV-1 sequence variation and replication capacity . . .	23
4.1.2	Description of Segal et al. (2004) HIV-1 dataset	23
4.2	Multiple testing procedures	24
4.3	Software implementation in SAS	25
4.4	Results	26



COBRA
A BEPRESS REPOSITORY

Collection of Biostatistics
Research Archive

1 Introduction

1.1 Motivation

Current statistical inference problems in areas such as genomics, astronomy, and marketing routinely involve the simultaneous test of thousands, or even millions, of null hypotheses. Examples of testing problems in genomics include:

- the identification of differentially expressed genes in microarray experiments, i.e., genes whose expression measures are associated with possibly censored responses or covariates;
- tests of association between gene expression measures and Gene Ontology (GO) annotation (www.geneontology.org);
- the identification of transcription factor binding sites in ChIP-Chip experiments, where chromatin immunoprecipitation (ChIP) of transcription factor bound DNA is followed by microarray hybridization (Chip) of the IP-enriched DNA (Keleş et al., 2004);
- the genetic mapping of complex traits using single nucleotide polymorphisms (SNP).

The above testing problems share the following general characteristics:

- inference for high-dimensional multivariate distributions, with complex and unknown dependence structures among variables;
- broad range of parameters of interest, such as, regression coefficients in model relating patient survival to genome-wide transcript levels or DNA copy numbers, pairwise gene correlations between transcript levels;
- many null hypotheses, in the thousands or even millions;
- complex dependence structures among test statistics, e.g., implied by the directed acyclic graph (DAG) structure of GO terms.

Motivated by these applications, we have developed and implemented (in R and SAS) resampling-based single-step and stepwise multiple testing

procedures (MTP) for controlling a broad class of Type I error rates, in testing problems involving general data generating distributions (with arbitrary dependence structures among variables), null hypotheses (defined in terms of submodels for the data generating distribution), and test statistics (e.g., t -statistics, F -statistics). The different components of our multiple testing methodology are treated in detail in the collection of related articles summarized below.

The early article of Pollard and van der Laan (2004) and subsequent article of Dudoit et al. (2004b) establish a general statistical framework for multiple hypothesis testing. A key feature of the proposed MTPs is the *test statistics null distribution* (rather than data generating null distribution) used to derive rejection regions (i.e., cut-offs) for the test statistics and resulting confidence regions and adjusted p -values. For Type I error rates defined as arbitrary parameters $\theta(F_{V_n})$ of the distribution of the number of Type I errors V_n (e.g., generalized family-wise error rate, $gFWER(k)$, or chance of at least $(k + 1)$ false positives), this null distribution is the asymptotic distribution of the vector of null value shifted and scaled test statistics. Resampling procedures (e.g., based on the non-parametric or model-based bootstrap) are proposed to conveniently obtain consistent estimators of the null distribution and the corresponding test statistic cut-offs and adjusted p -values (Dudoit and van der Laan, 2005; Dudoit et al., 2004b; van der Laan et al., 2004b; Pollard and van der Laan, 2004).

Pollard and van der Laan (2004) and Dudoit et al. (2004b) also derive *single-step common-cut-off and common-quantile procedures* for controlling general Type I error rates of the form $\theta(F_{V_n})$.

van der Laan et al. (2004b) focus on control of the family-wise error rate, $FWER = gFWER(0)$, and provide *step-down common-cut-off and common-quantile procedures*, based on maxima of test statistics (maxT) and minima of unadjusted p -values (minP), respectively.

van der Laan et al. (2004a), and subsequently Dudoit and van der Laan (2005) and Dudoit et al. (2004a), propose general classes of *augmentation multiple testing procedures* (AMTP), obtained by adding suitably chosen null hypotheses to the set of null hypotheses already rejected by an initial MTP. In particular, given *any* FWER-controlling procedure, they show how one can trivially obtain augmentation procedures controlling tail probabilities for the number (gFWER) and proportion (TPPFP) of false positives among the rejected hypotheses. The results of a simulation study comparing augmentation procedures to existing gFWER- and TPPFP-controlling MTPs

are reported in Dudoit et al. (2004a).

The software implementation of the aforementioned MTPs in the R package `multtest` is discussed in Pollard et al. (2005).

Finally, the multiple testing methodology and applications to genomic data analysis are the subject of a book in preparation for Springer (Dudoit and van der Laan, 2005).

1.2 Outline

Section 2 provides a summary of our proposed multiple testing procedures. Section 3 discusses their software implementation in a collection of SAS (Version 9) macros. Specifically, macros are provided for the bootstrap estimation of the test statistics null distribution, the FWER-controlling single-step maxT procedure, and gFWER- and TPPFP-controlling augmentation procedures (full code available in the Appendix). Finally, Section 4 describes the application of these MTPs to the HIV-1 sequence dataset of Segal et al. (2004), with the aim of relating codon genotypes in the protease and reverse transcriptase regions to viral replication capacity.

2 Multiple hypothesis testing methodology

2.1 Multiple hypothesis testing framework

Hypothesis testing is concerned with using observed data to test hypotheses, i.e., make decisions, regarding properties of the unknown data generating distribution. Below, we discuss in turn the main ingredients of a multiple testing problem, namely: data, null and alternative hypotheses, test statistics, multiple testing procedure, Type I and Type II errors, adjusted p -values, test statistics null distribution, rejection regions. Further detail on each of these components can be found in Dudoit and van der Laan (2005) and Dudoit et al. (2004b); specific proposals of MTPs are given in Sections 2.4 – 2.6.

Data. Let X_1, \dots, X_n be a *random sample* of n independent and identically distributed (i.i.d.) random variables, $X \sim P \in \mathcal{M}$, where the *data generating distribution* P is an element of a particular *statistical model* \mathcal{M} (i.e., a set of possibly non-parametric distributions).

Null and alternative hypotheses. In order to cover a broad class of testing problems, define M null hypotheses in terms of a collection of *submodels*, $\mathcal{M}(m) \subseteq \mathcal{M}$, $m = 1, \dots, M$, for the data generating distribution P . The M *null hypotheses* are defined as $H_0(m) \equiv \mathbb{I}(P \in \mathcal{M}(m))$ and the corresponding *alternative hypotheses* as $H_1(m) \equiv \mathbb{I}(P \notin \mathcal{M}(m))$.

In many testing problems, the submodels concern *parameters*, i.e., functions of the data generating distribution P , $\Psi(P) = \psi = (\psi(m) : m = 1, \dots, M)$, such as means, differences in means, correlation coefficients, and regression parameters in linear models, generalized linear models, survival models, time-series models, dose-response models, etc. One distinguishes between two types of testing problems: *one-sided tests*, where $H_0(m) = \mathbb{I}(\psi(m) \leq \psi_0(m))$, and *two-sided tests*, where $H_0(m) = \mathbb{I}(\psi(m) = \psi_0(m))$. The user-supplied hypothesized *null values*, $\psi_0(m)$, are frequently zero.

Let $\mathcal{H}_0 = \mathcal{H}_0(P) \equiv \{m : H_0(m) = 1\} = \{m : P \in \mathcal{M}(m)\}$ be the set of $h_0 \equiv |\mathcal{H}_0|$ true null hypotheses, where we note that \mathcal{H}_0 depends on the data generating distribution P . Let $\mathcal{H}_1 = \mathcal{H}_1(P) \equiv \mathcal{H}_0^c(P) = \{m : H_1(m) = 1\} = \{m : P \notin \mathcal{M}(m)\}$ be the set of $h_1 \equiv |\mathcal{H}_1| = M - h_0$ false null hypotheses, i.e., true positives. The goal of a multiple testing procedure is to accurately estimate the set \mathcal{H}_0 , and thus its complement \mathcal{H}_1 , while controlling probabilistically the number of false positives.

Test statistics. A testing procedure is a *data-driven* rule for deciding whether or not to *reject* each of the M null hypotheses $H_0(m)$, i.e., declare that $H_0(m)$ is false (zero) and hence $P \notin \mathcal{M}(m)$. The decisions to reject or not the null hypotheses are based on an M -vector of *test statistics*, $T_n = (T_n(m) : m = 1, \dots, M)$, that are functions $T_n(m) = T(X_1, \dots, X_n)(m)$ of the data, X_1, \dots, X_n . Denote the typically unknown (finite sample) *joint distribution* of the test statistics T_n by $Q_n = Q_n(P)$.

Single-parameter null hypotheses are commonly tested using *t-statistics*, i.e., standardized differences,

$$T_n(m) \equiv \frac{\text{Estimator} - \text{Null value}}{\text{Standard error}} = \sqrt{n} \frac{\psi_n(m) - \psi_0(m)}{\sigma_n(m)}. \quad (1)$$

In general, the M -vector $\psi_n = (\psi_n(m) : m = 1, \dots, M)$ denotes an asymptotically linear *estimator* of the parameter M -vector $\psi = (\psi(m) : m = 1, \dots, M)$ and $(\sigma_n(m)/\sqrt{n} : m = 1, \dots, M)$ denote consistent estimators of the *standard errors* of the components of ψ_n . For tests of means, one recovers the usual one-sample and two-sample *t-statistics*, where $\psi_n(m)$ and $\sigma_n(m)$

are based on empirical means and variances, respectively (e.g., two-sample t -statistic in Equation (24), p. 24, for the HIV-1 sequence data analysis of Section 4). In some settings, it may be appropriate to use (unstandardized) *difference statistics*, $T_n(m) \equiv \sqrt{n}(\psi_n(m) - \psi_0(m))$ (Pollard and van der Laan, 2004). Test statistics for other types of null hypotheses include F -statistics, χ^2 -statistics, and likelihood ratio statistics.

Multiple testing procedure. A *multiple testing procedure* (MTP) provides *rejection regions*, $\mathcal{C}_n(m)$, i.e., sets of values for each test statistic $T_n(m)$ that lead to the decision to reject the null hypothesis $H_0(m)$. In other words, a MTP produces a random (i.e., data-dependent) subset \mathcal{R}_n of rejected hypotheses that estimates \mathcal{H}_1 , the set of true positives,

$$\mathcal{R}_n = \mathcal{R}(T_n, Q_{0n}, \alpha) \equiv \{m : H_0(m) \text{ is rejected}\} = \{m : T_n(m) \in \mathcal{C}_n(m)\}, \quad (2)$$

where $\mathcal{C}_n(m) = \mathcal{C}(T_n, Q_{0n}, \alpha)(m)$, $m = 1, \dots, M$, denote possibly random rejection regions. The long notation $\mathcal{R}(T_n, Q_{0n}, \alpha)$ and $\mathcal{C}(T_n, Q_{0n}, \alpha)(m)$ emphasizes that the MTP depends on: (i) the *data*, X_1, \dots, X_n , through the M -vector of *test statistics*, $T_n = (T_n(m) : m = 1, \dots, M)$; (ii) a (estimated) test statistics *null distribution*, Q_{0n} , for deriving rejection regions for each $T_n(m)$ and the resulting adjusted p -values (Section 2.2); and (iii) the *nominal level* α , i.e., the desired upper bound for a suitably defined Type I error rate.

Unless specified otherwise, it is assumed that large values of the test statistic $T_n(m)$ provide evidence against the corresponding null hypothesis $H_0(m)$, that is, we consider rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), \infty)$, where $c_n(m)$ are to-be-determined *critical values*, or *cut-offs*, computed under the null distribution Q_{0n} for the test statistics T_n (Section 2.3).

Example: HIV-1 dataset. Suppose that, as in the analysis of the HIV-1 dataset of Section 4, one is interested in identifying codons in the protease (PR) and reverse transcriptase (RT) regions that are significantly associated with viral replication capacity (RC). The following data were collected on each of $n = 317$ patients: a replication capacity outcome/phenotype Y and an $M = 282$ -dimensional covariate vector $X = (X(m) : m = 1, \dots, M)$, of binary codon genotypes in the PR and RT regions (**zero** for wild-type codon and **one** for mutant codon). For the m th codon (i.e., m th hypothesis), the parameter of interest is the difference $\psi(m)$ in mean replication capacity of

viruses with mutant and wild-type codons, that is, $\psi(m) \equiv E[Y|X(m) = 1] - E[Y|X(m) = 0]$, $m = 1, \dots, M$. To identify codons that are associated with viral replication capacity, one can perform two-sided tests of the null hypotheses $H_0(m) = I(\psi(m) = 0)$ of no mean difference vs. the alternative hypotheses $H_1(m) = I(\psi(m) \neq 0)$, using pooled-variance two-sample t -statistics $T_n(m)$ (Equation (24), p. 24). The null hypotheses are rejected, i.e., the corresponding codon positions are declared significantly associated with replication capacity, for large absolute values of the test statistics $T_n(m)$.

Type I and Type II errors. In any testing situation, two types of errors can be committed: a *false positive*, or *Type I error*, is committed by rejecting a true null hypothesis, and a *false negative*, or *Type II error*, is committed when the test procedure fails to reject a false null hypothesis. The situation can be summarized by Table 1, below, where the number of Type I errors is $V_n \equiv |\mathcal{R}_n \cap \mathcal{H}_0| = \sum_{m \in \mathcal{H}_0} I(T_n(m) \in \mathcal{C}_n(m))$ and the number of Type II errors is $U_n \equiv |\mathcal{R}_n^c \cap \mathcal{H}_1| = \sum_{m \in \mathcal{H}_1} I(T_n(m) \notin \mathcal{C}_n(m))$. Note that both U_n and V_n depend on the unknown data generating distribution P through the unknown set of true null hypotheses $\mathcal{H}_0 = \mathcal{H}_0(P)$. The numbers $h_0 = |\mathcal{H}_0|$ and $h_1 = |\mathcal{H}_1| = M - h_0$ of true and false null hypotheses are *unknown parameters*, the number of rejected hypotheses $R_n \equiv |\mathcal{R}_n| = \sum_{m=1}^M I(T_n(m) \in \mathcal{C}_n(m))$ is an *observable random variable*, and the entries in the body of the table, U_n , $h_1 - U_n$, V_n , and $h_0 - V_n$, are *unobservable random variables* (depending on P through $\mathcal{H}_0(P)$).

Ideally, one would like to simultaneously minimize both the number of Type I errors and the number of Type II errors. Unfortunately, this is not feasible and one seeks a *trade-off* between the two types of errors. A standard approach is to specify an acceptable level α for the Type I error rate and derive testing procedures, i.e., rejection regions, that aim to minimize the Type II error rate, i.e., maximize *power*, within the class of procedures with Type I error rate at most α .

Type I error rates. When testing multiple hypotheses, there are many possible definitions for the Type I error rate (and power) of a test procedure. Accordingly, we adopt the general framework proposed in Dudoit and van der Laan (2005) and Dudoit et al. (2004b), and define Type I error rates as *parameters*, $\theta_n = \theta(F_{V_n, R_n})$, of the joint distribution F_{V_n, R_n} of the numbers of Type I errors V_n and rejected hypotheses R_n . Such a general representation covers the following commonly-used Type I error rates.

Generalized family-wise error rate (gFWER), or probability of at least $(k + 1)$ Type I errors, $k = 0, \dots, (h_0 - 1)$,

$$gFWER(k) \equiv Pr(V_n > k) = 1 - F_{V_n}(k), \quad (3)$$

where F_{V_n} is the discrete cumulative distribution function (c.d.f.) on $\{0, \dots, M\}$ for the number of Type I errors, V_n . When $k = 0$, the gFWER is the usual *family-wise error rate* (FWER), or probability of at least one Type I error,

$$FWER \equiv Pr(V_n > 0) = 1 - F_{V_n}(0). \quad (4)$$

The FWER is controlled, in particular, by the classical Bonferroni procedure.

Per-comparison error rate (PCER), or expected value of the proportion of Type I errors among the M tests,

$$PCER \equiv \frac{1}{M} E[V_n] = \frac{1}{M} \int v dF_{V_n}(v). \quad (5)$$

Tail probabilities for the proportion of false positives (TPPFP) among the rejected hypotheses,

$$TPPFP(q) \equiv Pr(V_n/R_n > q) = 1 - F_{V_n/R_n}(q), \quad q \in (0, 1), \quad (6)$$

where F_{V_n/R_n} is the c.d.f. for the proportion V_n/R_n of false positives among the rejected hypotheses, with the convention that $V_n/R_n \equiv 0$ if $R_n = 0$.

False discovery rate (FDR), or expected value of the proportion of false positives among the rejected hypotheses,

$$FDR \equiv E[V_n/R_n] = \int q dF_{V_n/R_n}(q), \quad (7)$$

again with the convention that $V_n/R_n \equiv 0$ if $R_n = 0$ (Benjamini and Hochberg, 1995).

Note that while the gFWER is a parameter of only the *marginal* distribution F_{V_n} of the number of Type I errors V_n (tail probability, or survivor

function, for V_n), the TPPFP is a parameter of the *joint* distribution of (V_n, R_n) (tail probability, or survivor function, for V_n/R_n). Error rates based on the *proportion* of false positives (e.g., TPPFP and FDR) are especially appealing for large-scale testing problems such as those encountered in genomics, compared to error rates based on the *number* of false positives (e.g., gFWER), as they do not increase exponentially with the number of tested hypotheses.

The aforementioned error rates are part of the broad class of Type I error rates considered in Dudoit and van der Laan (2005) and Dudoit et al. (2004a), and defined as tail probabilities $Pr(g(V_n, R_n) > q)$ and expected values $E[g(V_n, R_n)]$ for an arbitrary function $g(V_n, R_n)$ of the numbers of false positives V_n and rejected hypotheses R_n . The gFWER and TPPFP correspond to the special cases $g(V_n, R_n) = V_n$ and $g(V_n, R_n) = V_n/R_n$, respectively.

Adjusted p -values. The notion of p -value extends directly to multiple testing problems, as follows. Given a MTP $\mathcal{R}_n(\alpha) = \mathcal{R}(T_n, Q_{0n}, \alpha)$, the *adjusted p -value* $\tilde{P}_{0n}(m) = \tilde{P}(T_n, Q_{0n})(m)$, for null hypothesis $H_0(m)$, is defined as the smallest Type I error level α at which one would reject $H_0(m)$, that is,

$$\begin{aligned} \tilde{P}_{0n}(m) &\equiv \inf \{ \alpha \in [0, 1] : m \in \mathcal{R}_n(\alpha) \} \\ &= \inf \{ \alpha \in [0, 1] : T_n(m) \in \mathcal{C}_n(m) \}, \quad m = 1, \dots, M. \end{aligned} \quad (8)$$

Note that *unadjusted* or *marginal p -values*, for the test of a single hypothesis, correspond to the special case $M = 1$. For a continuous null distribution Q_{0n} , the unadjusted p -value for null hypothesis $H_0(m)$ is given by $P_{0n}(m) = P(T_n(m), Q_{0n,m}) = \bar{Q}_{0n,m}(T_n(m))$, where $Q_{0n,m}$ and $\bar{Q}_{0n,m}$ denote, respectively, the marginal c.d.f.'s and survivor functions for Q_{0n} . For example, the adjusted p -values for the classical Bonferroni procedure for FWER control are given by $\tilde{P}_{0n}(m) = \min(MP_{0n}(m), 1)$.

As in single hypothesis tests, the smaller the adjusted p -value, the stronger the evidence against the corresponding null hypothesis. If $\mathcal{R}_n(\alpha)$ is right-continuous at α , in the sense that $\lim_{\alpha' \downarrow \alpha} \mathcal{R}_n(\alpha') = \mathcal{R}_n(\alpha)$, then one has two equivalent representations for the MTP, in terms of rejection regions for the test statistics and in terms of adjusted p -values,

$$\mathcal{R}_n(\alpha) = \{ m : T_n(m) \in \mathcal{C}_n(m) \} = \{ m : \tilde{P}_{0n}(m) \leq \alpha \}. \quad (9)$$

Reporting the results of a MTP in terms of adjusted p -values, as opposed

to only rejection or not of the hypotheses, offers several advantages. (i) Adjusted p -values can be defined for *any Type I error rate* (gFWER, TPPFP, FDR, etc.). (ii) They reflect the strength of the evidence against each null hypothesis in terms of the Type I error rate for the *entire MTP*. (iii) They are *flexible summaries* of a MTP, in that results are supplied for *all levels* α , i.e., the level α need not be chosen ahead of time. (iv) Finally, adjusted p -values provide convenient benchmarks to *compare* different MTPs, whereby smaller adjusted p -values indicate a less conservative procedure.

Confidence regions. For the test of single-parameter null hypotheses and for any Type I error rate of the form $\theta(F_{V_n})$, Dudoit and van der Laan (2005) and Pollard and van der Laan (2004) provide results on the correspondence between single-step MTPs and θ -specific *confidence regions*.

2.2 Test statistics null distribution

2.2.1 Characterization of the test statistics null distribution

One of the main tasks in specifying a MTP is to derive rejection regions for the test statistics such that the Type I error rate is controlled at a desired level α , i.e., such that $\theta(F_{V_n, R_n}) \leq \alpha$, for *finite sample control*, or $\limsup_n \theta(F_{V_n, R_n}) \leq \alpha$, for *asymptotic control*. It is common practice, especially for FWER control, to set $\alpha = 0.05$. However, one is immediately faced with the problem that the *true distribution* $Q_n = Q_n(P)$ of the test statistics T_n is usually *unknown*, and hence, so are the distributions of the numbers of Type I errors, $V_n = \sum_{m \in \mathcal{H}_0} \mathbf{I}(T_n(m) \in \mathcal{C}_n(m))$, and rejected hypotheses, $R_n = \sum_{m=1}^M \mathbf{I}(T_n(m) \in \mathcal{C}_n(m))$. In practice, the test statistics *true distribution* $Q_n(P)$ is replaced by a *null distribution* Q_0 (or estimate thereof, Q_{0n}), in order to derive rejection regions and resulting adjusted p -values.

The choice of null distribution Q_0 is crucial, in order to ensure that (finite sample or asymptotic) control of the Type I error rate under the *assumed* null distribution Q_0 does indeed provide the required control under the *true* distribution $Q_n(P)$. For proper control, the null distribution Q_0 must be such that the Type I error rate under this assumed null distribution *dominates* the Type I error rate under the true distribution $Q_n(P)$. That is, one must have $\theta(F_{V_n, R_n}) \leq \theta(F_{V_0, R_0})$, for finite sample control, and $\limsup_n \theta(F_{V_n, R_n}) \leq \theta(F_{V_0, R_0})$, for asymptotic control, where V_0 and R_0 denote, respectively, the numbers of Type I errors and rejected hypotheses under the assumed null

distribution Q_0 .

For error rates $\theta(F_{V_n})$, defined as arbitrary parameters of the distribution of the number of Type I errors V_n , we propose as null distribution the asymptotic distribution $Q_0 = Q_0(P)$ of the M -vector Z_n of null value shifted and scaled test statistics (Dudoit and van der Laan, 2005; Dudoit et al., 2004b; van der Laan et al., 2004b; Pollard and van der Laan, 2004),

$$Z_n(m) \equiv \sqrt{\min\left(1, \frac{\tau_0(m)}{\text{Var}[T_n(m)]}\right)} \left(T_n(m) + \lambda_0(m) - E[T_n(m)]\right). \quad (10)$$

For the test of single-parameter null hypotheses using t -statistics, the null values are $\lambda_0(m) = 0$ and $\tau_0(m) = 1$. For testing the equality of K population means using F -statistics, the null values are $\lambda_0(m) = 1$ and $\tau_0(m) = 2/(K - 1)$, under the assumption of equal variances in the different populations. Dudoit et al. (2004b) and van der Laan et al. (2004b) prove that this null distribution does indeed provide the desired asymptotic control of the Type I error rate $\theta(F_{V_n})$, for general data generating distributions (with arbitrary dependence structures among variables), null hypotheses (defined in terms of submodels for the data generating distribution), and test statistics (e.g., t -statistics, F -statistics).

For a broad class of testing problems, such as the test of single-parameter null hypotheses using t -statistics (as in Equation (1)), the null distribution Q_0 is an M -variate Gaussian distribution with mean vector zero and covariance matrix $\Sigma^*(P)$: $Q_0 = Q_0(P) \equiv N(0, \Sigma^*(P))$. For tests of means, where the parameter of interest is the M -dimensional mean vector $\Psi(P) = \psi = E[X]$, the estimator ψ_n is simply the M -vector of sample averages and $\Sigma^*(P)$ is the correlation matrix of $X \sim P$, $\text{Cor}[X]$. More generally, for an asymptotically linear estimator ψ_n , $\Sigma^*(P)$ is the correlation matrix of the vector influence curve (IC).

Note that the following important points distinguish our approach from existing approaches to Type I error rate control. Firstly, we are only concerned with Type I error control under the *true data generating distribution* P . The notions of weak and strong control (and associated subset pivotality, Westfall and Young (1993), p. 42–43) are therefore irrelevant to our approach. Secondly, we propose a *null distribution for the test statistics* (i.e., $T_n \sim Q_0$), and not a data generating null distribution (i.e., $X \sim P_0 \in \bigcap_{m=1}^M \mathcal{M}(m)$). The latter practice does not necessarily provide proper Type I error control, as the test statistics' *assumed* null distribution $Q_n(P_0)$ and

their *true* distribution $Q_n(P)$ may have different dependence structures (in the limit) for the true null hypotheses \mathcal{H}_0 .

2.2.2 Bootstrap estimation of the test statistics null distribution

In practice, since the data generating distribution P is unknown, then so is the proposed null distribution $Q_0 = Q_0(P)$. Resampling procedures, such as bootstrap Procedure 1, below, may be used to conveniently obtain consistent estimators Q_{0n} of the null distribution Q_0 and of the corresponding test statistic cut-offs and adjusted p -values. The reader is referred to our earlier articles and a book in preparation for further detail on the choice of test statistics T_n , null distribution Q_0 , and approaches for estimating this null distribution (Dudoit and van der Laan, 2005; Dudoit et al., 2004b; van der Laan et al., 2004b; Pollard and van der Laan, 2004). Accordingly, we take the test statistics T_n and their null distribution Q_0 (or estimate thereof, Q_{0n}) as given, and denote the set and number of rejected hypotheses by $\mathcal{R}_n(\alpha) = \mathcal{R}(T_n, Q_{0n}, \alpha)$ and $R_n(\alpha)$ (or the shorter \mathcal{R}_n and R_n), respectively, to emphasize only the dependence on the nominal Type I error level α .



Procedure 1 [Bootstrap estimation of the null distribution Q_0]

1. Let P_n^* denote an estimator of the data generating distribution P . For the non-parametric bootstrap, P_n^* is simply the empirical distribution P_n , that is, samples of size n are drawn at random, with replacement from the observed data, X_1, \dots, X_n . For the model-based bootstrap, P_n^* is based on a model \mathcal{M} for the data generating distribution P , such as the family of M -variate Gaussian distributions.
2. Generate B bootstrap samples, each consisting of n i.i.d. realizations of a random variable $X^\# \sim P_n^*$.
3. For the b th bootstrap sample, $b = 1, \dots, B$, compute an M -vector of test statistics, $T_n^\#(\cdot, b) = (T_n^\#(m, b) : m = 1, \dots, M)$. Arrange these bootstrap statistics in an $M \times B$ matrix, $\mathbf{T}_n^\# = (T_n^\#(m, b))$, with rows corresponding to the M null hypotheses and columns to the B bootstrap samples.
4. Compute row means, $E[T_n^\#(m, \cdot)]$, and row variances, $\text{Var}[T_n^\#(m, \cdot)]$, of the matrix $\mathbf{T}_n^\#$, to yield estimates of the true means $E[T_n(m)]$ and variances $\text{Var}[T_n(m)]$ of the test statistics, respectively.
5. Obtain an $M \times B$ matrix, $\mathbf{Z}_n^\# = (Z_n^\#(m, b))$, of null value shifted and scaled bootstrap statistics $Z_n^\#(m, b)$, by row-shifting and scaling the matrix $\mathbf{T}_n^\#$ using the bootstrap estimates of $E[T_n(m)]$ and $\text{Var}[T_n(m)]$ and the user-supplied null values $\lambda_0(m)$ and $\tau_0(m)$. That is, compute

$$Z_n^\#(m, b) \equiv \sqrt{\min\left(1, \frac{\tau_0(m)}{\text{Var}[T_n^\#(m, \cdot)]}\right)} \times \left(T_n^\#(m, b) + \lambda_0(m) - E[T_n^\#(m, \cdot)]\right). \quad (11)$$

6. The bootstrap estimate Q_{0n} of the null distribution Q_0 is the empirical distribution of the B columns $Z_n^\#(\cdot, b)$ of matrix $\mathbf{Z}_n^\#$.

2.3 Rejection regions

Having selected a suitable test statistics null distribution, there remains the main task of specifying rejection regions for each null hypothesis, i.e., cut-offs for each test statistic. Among the different approaches for defining rejection regions, we distinguish between the following.

Common-cut-off vs. common-quantile multiple testing procedures.

In *common-cut-off procedures*, the same cut-off c_0 is used for each test statistic (cf. FWER-controlling maxT Procedures 2 and 4, based on maxima of test statistics). In contrast, in *common-quantile procedures*, the cut-offs are the δ_0 -quantiles of the marginal null distributions of the test statistics (cf. FWER-controlling minP Procedures 3 and 5, based on minima of unadjusted p -values). The latter procedures tend to be more “balanced” than the former, as the transformation to p -values places the null hypotheses on an equal footing. However, this comes at the expense of increased computational complexity.

Single-step vs. stepwise multiple testing procedures. In *single-step procedures*, each null hypothesis is evaluated using a rejection region that is independent of the results of the tests of other hypotheses. Improvement in power, while preserving Type I error rate control, may be achieved by *stepwise procedures*, in which rejection of a particular null hypothesis depends on the outcome of the tests of other hypotheses. That is, the (single-step) test procedure is applied to a sequence of successively smaller nested random (i.e., data-dependent) subsets of null hypotheses, defined by the ordering of the test statistics (common-cut-off MTPs) or unadjusted p -values (common-quantile MTPs). In *step-down procedures*, the hypotheses corresponding to the *most significant* test statistics (i.e., largest absolute test statistics or smallest unadjusted p -values) are considered successively, with further tests depending on the outcome of earlier ones. As soon as one fails to reject a null hypothesis, no further hypotheses are rejected. In contrast, for *step-up procedures*, the hypotheses corresponding to the *least significant* test statistics are considered successively, again with further tests depending on the outcome of earlier ones. As soon as one hypothesis is rejected, all remaining more significant hypotheses are rejected.

Marginal vs. joint multiple testing procedures. *Marginal multiple testing procedures* are based solely on the marginal distributions

of the test statistics, i.e., on cut-off rules for individual test statistics or their corresponding unadjusted p -values (e.g., classical Bonferroni FWER-controlling single-step procedure). In contrast, *joint multiple testing procedures* take into account the dependence structure of the test statistics (e.g., gFWER-controlling single-step common-cut-off and common-quantile Procedures 2 and 3, based on maxima of test statistics and minima of unadjusted p -values, respectively).

The next three sections summarize three general approaches for deriving rejection regions and corresponding adjusted p -values. The main steps in applying a multiple testing procedure are listed in the flowchart of Table 2.

Single-step common-cut-off and common-quantile procedures for controlling general Type I error rates $\theta(F_{V_n})$: Procedures 2 and 3, Section 2.4; details in Dudoit and van der Laan (2005), Dudoit et al. (2004b), and Pollard and van der Laan (2004).

Step-down common-cut-off (maxT) and common-quantile (minP) procedures for controlling the FWER: Procedures 4 and 5, Section 2.5; details in Dudoit and van der Laan (2005) and van der Laan et al. (2004b).

Augmentation procedures for controlling the gFWER and TPPFP, based on an initial FWER-controlling procedure: Procedures 6 and 7, Section 2.6; details and extensions in Dudoit and van der Laan (2005), Dudoit et al. (2004a), and van der Laan et al. (2004a).

2.4 Single-step procedures for controlling general Type I error rates $\theta(F_{V_n})$

Dudoit et al. (2004b) and Pollard and van der Laan (2004) propose single-step common-cut-off and common-quantile procedures for controlling arbitrary parameters $\theta(F_{V_n})$ of the distribution of the number of Type I errors. The main idea is to substitute control of the parameter $\theta(F_{V_n})$, for the *unknown, true distribution* F_{V_n} of the number of Type I errors, by control of the corresponding parameter $\theta(F_{R_0})$, for the *known, null distribution* F_{R_0} of the number of rejected hypotheses. That is, one considers single-step procedures of the form $\mathcal{R}_n(\alpha) \equiv \{m : T_n(m) > c_n(m)\}$, where the cut-offs $c_n(m)$ are chosen so that $\theta(F_{R_0}) \leq \alpha$, for $R_0 \equiv \sum_{m=1}^M \mathbf{I}(Z(m) > c_n(m))$ and $Z \sim Q_0$.

Among the class of MTPs that satisfy $\theta(F_{R_0}) \leq \alpha$, Dudoit et al. (2004b) and Pollard and van der Laan (2004) propose two procedures, based on common cut-offs and common-quantile cut-offs, respectively (Procedures 2 and 1 of Dudoit et al. (2004b)). The procedures are summarized below and the reader is referred to the articles for proofs and details on the derivation of cut-offs and adjusted p -values.

Procedure 2 [General θ -controlling single-step common-cut-off procedure] *The set of rejected hypotheses for the general θ -controlling single-step common-cut-off procedure is of the form $\mathcal{R}_n(\alpha) \equiv \{m : T_n(m) > c_0\}$, where the common cut-off c_0 is the smallest (i.e., least conservative) value for which $\theta(F_{R_0}) \leq \alpha$.*

For $gFWER(k)$ control (i.e., $\theta(F_{V_n}) = 1 - F_{V_n}(k)$), the procedure is based on the $(k + 1)$ st ordered test statistic. The adjusted p -values for the single-step $T(k + 1)$ procedure are given by

$$\tilde{p}_{0n}(m) = Pr_{Q_0}(Z^\circ(k + 1) \geq t_n(m)), \quad m = 1, \dots, M, \quad (12)$$

where $Z^\circ(m)$ denotes the m th ordered component of $Z = (Z(m) : m = 1, \dots, M) \sim Q_0$, so that $Z^\circ(1) \geq \dots \geq Z^\circ(M)$.

For $FWER$ control ($k = 0$), one recovers the single-step $\max T$ procedure, based on the maximum test statistic, $Z^\circ(1) = \max_m Z(m)$, with adjusted p -values given by

$$\tilde{p}_{0n}(m) = Pr_{Q_0} \left(\max_{m \in \{1, \dots, M\}} Z(m) \geq t_n(m) \right), \quad m = 1, \dots, M. \quad (13)$$

Procedure 3 [General θ -controlling single-step common-quantile procedure] *The set of rejected hypotheses for the general θ -controlling single-step common-quantile procedure is of the form $\mathcal{R}_n(\alpha) \equiv \{m : T_n(m) > c_0(m)\}$, where $c_0(m) = Q_{0,m}^{-1}(\delta_0)$ is the δ_0 -quantile of the marginal null distribution $Q_{0,m}$ of the test statistic for the m th null hypothesis, i.e., the smallest value c such that $Q_{0,m}(c) = Pr_{Q_0}(Z(m) \leq c) \geq \delta_0$ for $Z \sim Q_0$. Here, δ_0 is chosen as the smallest (i.e., least conservative) value for which $\theta(F_{R_0}) \leq \alpha$.*

For $gFWER(k)$ control (i.e., $\theta(F_{V_n}) = 1 - F_{V_n}(k)$), the procedure is based on the $(k + 1)$ st ordered unadjusted p -value. Specifically, let $\bar{Q}_{0,m} \equiv 1 - Q_{0,m}$ denote the survivor functions for the marginal null distributions $Q_{0,m}$ and define unadjusted p -values $P_0(m) \equiv \bar{Q}_{0,m}(Z(m))$ and $P_{0n}(m) \equiv \bar{Q}_{0,m}(T_n(m))$,

for $Z \sim Q_0$ and $T_n \sim Q_n$, respectively. The adjusted p -values for the single-step $P(k+1)$ procedure are given by

$$\tilde{p}_{0n}(m) = Pr_{Q_0}(P_0^\circ(k+1) \leq p_{0n}(m)), \quad m = 1, \dots, M, \quad (14)$$

where $P_0^\circ(m)$ denotes the m th ordered component of the M -vector of unadjusted p -values $P_0 = (P_0(m) : m = 1, \dots, M)$, so that $P_0^\circ(1) \leq \dots \leq P_0^\circ(M)$.

For FWER control ($k = 0$), one recovers the single-step minP procedure, based on the minimum unadjusted p -value, $P_0^\circ(1) = \min_m P_0(m)$, with adjusted p -values given by

$$\tilde{p}_{0n}(m) = Pr_{Q_0} \left(\min_{m \in \{1, \dots, M\}} P_0(m) \leq p_{0n}(m) \right), \quad m = 1, \dots, M. \quad (15)$$

2.5 Step-down procedures for controlling the family-wise error rate

van der Laan et al. (2004b) propose step-down common-cut-off (maxT) and common-quantile (minP) procedures for controlling the family-wise error rate, FWER. These procedures are similar in spirit to their single-step counterparts in Section 2.4, for the special case $\theta(F_{V_n}) = 1 - F_{V_n}(0)$, with the important step-down distinction that hypotheses are considered successively, from most significant to least significant, with further tests depending on the outcome of earlier ones. That is, the test procedure is applied to a sequence of successively smaller nested random (i.e., data-dependent) subsets of null hypotheses, defined by the ordering of the test statistics (common-cut-off MTPs) or unadjusted p -values (common-quantile MTPs).

Procedure 4 [FWER-controlling step-down common-cut-off (maxT) procedure] Let $O_n(m)$ denote the indices for the ordered test statistics $T_n(m)$, so that $T_n(O_n(1)) \geq \dots \geq T_n(O_n(M))$. The step-down common-cut-off (maxT) procedure is based on the distributions of maxima of test statistics over the nested subsets of ordered null hypotheses $\bar{O}_n(h) \equiv \{O_n(h), \dots, O_n(M)\}$. The adjusted p -values are given by

$$\tilde{p}_{0n}(o_n(m)) = \max_{h=1, \dots, m} \left\{ Pr_{Q_0} \left(\max_{l \in \bar{O}_n(h)} Z(l) \geq t_n(o_n(h)) \right) \right\}, \quad (16)$$

where $Z = (Z(m) : m = 1, \dots, M) \sim Q_0$.

Thus, unlike single-step maxT Procedure 2, based solely on the distribution of the maximum test statistic over all M hypotheses, the step-down common cut-offs and corresponding adjusted p -values are based on the distributions of maxima of test statistics over successively smaller nested random subsets of null hypotheses. Taking maxima of the probabilities over $h \in \{1, \dots, m\}$ enforces monotonicity of the adjusted p -values and ensures that the procedure is indeed step-down, that is, one can only reject a particular hypothesis provided all hypotheses with more significant (i.e., larger) test statistics were rejected beforehand.

Likewise, the step-down common-quantile cut-offs and corresponding adjusted p -values are based on the distributions of minima of unadjusted p -values over successively smaller nested random subsets of null hypotheses.

Procedure 5 [FWER-controlling step-down common-quantile (minP) procedure] Let $O_n(m)$ denote the indices for the ordered unadjusted p -values $P_{0n}(m)$, so that $P_{0n}(O_n(1)) \leq \dots \leq P_{0n}(O_n(M))$. The step-down common-quantile (minP) procedure is based on the distributions of minima of unadjusted p -values over the nested subsets of ordered null hypotheses $\bar{O}_n(h) \equiv \{O_n(h), \dots, O_n(M)\}$. The adjusted p -values are given by

$$\tilde{p}_{0n}(o_n(m)) = \max_{h=1, \dots, m} \left\{ Pr_{Q_0} \left(\min_{l \in \bar{O}_n(h)} P_0(l) \leq p_{0n}(o_n(h)) \right) \right\}, \quad (17)$$

where $P_0(m) \equiv \bar{Q}_{0,m}(Z(m))$ and $P_{0n}(m) \equiv \bar{Q}_{0,m}(T_n(m))$, for $Z \sim Q_0$ and $T_n \sim Q_n$, respectively.

2.6 Augmentation multiple testing procedures for controlling tail probability error rates

van der Laan et al. (2004a), and subsequently Dudoit and van der Laan (2005) and Dudoit et al. (2004a), propose *augmentation multiple testing procedures* (AMTP), obtained by adding suitably chosen null hypotheses to the set of null hypotheses already rejected by an initial MTP. Specifically, given *any* initial procedure controlling the generalized family-wise error rate, augmentation procedures are derived for controlling Type I error rates defined as tail probabilities and expected values for arbitrary functions $g(V_n, R_n)$ of the numbers of Type I errors and rejected hypotheses (e.g., proportion $g(V_n, R_n) = V_n/R_n$ of false positives among the rejected hypotheses). Adjusted p -values for the AMTP are shown to be simply shifted versions of the

adjusted p -values of the original MTP. The important practical implication of these results is that *any* FWER-controlling MTP and its corresponding adjusted p -values immediately provide multiple testing procedures controlling a broad class of Type I error rates and their adjusted p -values. One can therefore build on the large pool of available FWER-controlling procedures, such as the single-step and step-down maxT and minP procedures discussed in Sections 2.4 and 2.5, above.

Augmentation procedures for controlling tail probabilities of the number (gFWER) and proportion (TPFP) of false positives, based on an initial FWER-controlling procedure, are treated in detail in Dudoit et al. (2004a) and van der Laan et al. (2004a) and are summarized below. The gFWER and TPFP correspond to the special cases $g(V_n, R_n) = V_n$ and $g(V_n, R_n) = V_n/R_n$, respectively.

Denote the adjusted p -values for the initial FWER-controlling procedure by $\tilde{P}_{0n}(m)$. Order the M null hypotheses according to these p -values, from smallest to largest, that is, define indices $O_n(m)$, so that $\tilde{P}_{0n}(O_n(1)) \leq \dots \leq \tilde{P}_{0n}(O_n(M))$. Then, for a nominal level α test, the initial FWER-controlling procedure rejects the $R_n(\alpha)$ null hypotheses

$$\mathcal{R}_n(\alpha) \equiv \{m : \tilde{P}_{0n}(m) \leq \alpha\}. \quad (18)$$

Procedure 6 [gFWER-controlling augmentation multiple testing procedure] For control of $gFWER(k)$ at level α , given an initial FWER-controlling procedure, reject the $R_n(\alpha)$ hypotheses specified by this MTP, as well as the next $A_n(\alpha)$ most significant null hypotheses,

$$A_n(\alpha) = \min\{k, M - R_n(\alpha)\}. \quad (19)$$

The adjusted p -values $\tilde{P}_{0n}^+(O_n(m))$ for the new gFWER-controlling AMTP are simply k -shifted versions of the adjusted p -values of the initial FWER-controlling MTP, with the first k adjusted p -values set to zero. That is,

$$\tilde{P}_{0n}^+(O_n(m)) = \begin{cases} 0, & \text{if } m \leq k \\ \tilde{P}_{0n}(O_n(m - k)), & \text{if } m > k \end{cases}. \quad (20)$$

The AMTP thus guarantees at least k rejected hypotheses.

Procedure 7 [TPFP-controlling augmentation multiple testing procedure] For control of $TPFP(q)$ at level α , given an initial FWER-controlling

procedure, reject the $R_n(\alpha)$ hypotheses specified by this MTP, as well as the next $A_n(\alpha)$ most significant null hypotheses,

$$\begin{aligned} A_n(\alpha) &= \max \left\{ m \in \{0, \dots, M - R_n(\alpha)\} : \frac{m}{m + R_n(\alpha)} \leq q \right\} \\ &= \min \left\{ \left\lfloor \frac{qR_n(\alpha)}{1 - q} \right\rfloor, M - R_n(\alpha) \right\}, \end{aligned} \quad (21)$$

where the floor $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x , i.e., $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$. That is, keep rejecting null hypotheses until the ratio of additional rejections to the total number of rejections reaches the allowed proportion q of false positives. The adjusted p -values $\tilde{P}_{0n}^+(O_n(m))$ for the new TPPFP-controlling AMTP are simply mq -shifted versions of the adjusted p -values of the initial FWER-controlling MTP. That is,

$$\tilde{P}_{0n}^+(O_n(m)) = \tilde{P}_{0n}(O_n(\lceil (1 - q)m \rceil)), \quad m = 1, \dots, M, \quad (22)$$

where the ceiling $\lceil x \rceil$ denotes the least integer greater than or equal to x , i.e., $\lceil x \rceil - 1 < x \leq \lceil x \rceil$.

FDR-controlling procedures. Given any TPPFP-controlling procedure, van der Laan et al. (2004a) derive two simple (conservative) FDR-controlling procedures. The more general and conservative procedure controls the FDR at nominal level α , by controlling $TPPFP(\alpha/2)$ at level $\alpha/2$. The less conservative procedure controls the FDR at nominal level α , by controlling $TPPFP(1 - \sqrt{1 - \alpha})$ at level $1 - \sqrt{1 - \alpha}$. The reader is referred to the original article for details and proofs of FDR control (Section 2.4, Theorem 3).

3 Software implementation in SAS

This section discusses the software implementation in SAS (Version 9) of the following three MTPs: the FWER-controlling single-step maxT Procedure 2, the gFWER-controlling augmentation Procedure 6, and the TPPFP-controlling augmentation Procedure 7. The SAS macros described below were written to compute various components of a MTP, including test statistics T_n , bootstrap estimates Q_{0n} of the test statistics null distribution Q_0 (Procedure 1), and adjusted p -values $\tilde{P}_{0n}(m)$. The full code is provided in the Appendix and as a text file on the website www.stat.berkeley.edu/~sandrine/publications.html.

%lmt: The **%lmt** macro takes as input a SAS dataset of the form [y:X] (e.g., **resample.hivdata** for the HIV-1 sequence analysis of Section 4), with records corresponding to n observations and with the first column referring to an outcome Y and the remaining columns to an M -dimensional covariate vector $X = (X(m) : m = 1, \dots, M)$. This macro uses PROC REG to compute t -statistics $T_n(m)$, for the univariate linear regression of the outcome Y on each of the M covariates $X(m)$, $m = 1, \dots, M$. The test statistics T_n are stored in the dataset **tstats**.

%boot: The **%boot** macro generates B (macro variable **&boots**) bootstrap samples of the original dataset and computes M -vectors of test statistics $T_n^\#$ for each of these B bootstrapped datasets using the macro **%lmt**. Specifically, the rows of the dataset [y:X] are sampled at random, with replacement using PROC SURVEYSELECT. Note that PROC SURVEYSELECT uses the bootstrap iteration index as the seed and the **method=urs** command for “unrestricted random sampling”, i.e., sampling at random, with replacement. The bootstrap test statistics $T_n^\#$ are stored in the dataset **tstatsB**.

%bootnull: The **%bootnull** macro reads in the B bootstrap M -vectors of test statistics $T_n^\#$ produced by the **%boot** macro and stored in the dataset **tstatsB**. PROC IML is used to compute B M -vectors of corresponding null value shifted test statistics $Z_n^\#$, which are then stored in the dataset **Qo**. The empirical distribution of these new M -vectors of test statistics $Z_n^\#$ yields a bootstrap estimate Q_{0n} of the null distribution (Procedure 1).

%ssmaxT: The **%ssmaxT** macro takes as input the M -vector of test statistics T_n for the original data (i.e., dataset **tstats** from macro **%lmt**) and B M -vectors of null value shifted test statistics $Z_n^\#$, corresponding to a bootstrap estimate Q_{0n} of the null distribution (i.e., dataset **Qo** of dimension BM from macro **%bootnull**). For each of the B bootstrap samples, the maximum of the M test statistics is computed using PROC IML (i.e., row maxima of the $B \times M$ matrix obtained from the dataset **Qo**). Single-step maxT adjusted p -values are computed for each of the M hypotheses as the proportions of the B maxima that exceed the corresponding test statistics for the original dataset (Equation (13), Procedure 2). The adjusted p -values are stored in the dataset **fwer**. Note that the current implementation of **%ssmaxT** provides p -values

for two-sided tests only (i.e., based on the absolute values of the test statistics).

%gfwcr: Given an allowed number k of false positives and adjusted p -values for an arbitrary initial FWER-controlling MTP (e.g., dataset **fwer** from macro **%ssmaxT**), the **%gfwcr** macro uses PROC IML to compute adjusted p -values for the gFWER-controlling augmentation Procedure 6.

%tppfp: Given an allowed proportion q of false positives and adjusted p -values for an arbitrary initial FWER-controlling MTP (e.g., dataset **fwer** from macro **%ssmaxT**), the **%tppfp** macro uses PROC IML to compute adjusted p -values for the TPPFP-controlling augmentation Procedure 7.

Note that the macros **%fwer**, **%gfwcr**, and **%tppfp** return adjusted p -values as datasets, as opposed to matrices, to facilitate the identification and labeling of hypotheses (i.e., columns in the original dataset).

Our motivation for developing the above macros was the analysis of the HIV-1 dataset discussed in Section 4, below. The user could easily modify and extend this collection of macros to adapt to other data structures and/or testing problems.

4 Application to HIV-1 sequence data

In this section, we apply the multiple testing procedures implemented in the SAS macros of Section 3 to the HIV-1 dataset of Segal et al. (2004), with the aim of relating HIV-1 sequence variation to viral replication capacity. Specifically, multiple testing procedures are applied to identify codons which are significantly associated with viral replication capacity, based on t -statistics for the univariate linear regression of the replication capacity phenotype on individual codon genotypes. Brief descriptions of the HIV-1 dataset, multiple testing procedures, software implementation in SAS, and results are given next.

4.1 HIV-1 dataset

4.1.1 HIV-1 sequence variation and replication capacity

Studying sequence variation for the Human Immunodeficiency Virus Type 1 (HIV-1) genome could potentially give important insight into *genotype-phenotype associations* for the Acquired Immune Deficiency Syndrome (AIDS).

In this context, a key *phenotype* is the replication capacity (RC) of HIV-1, as it reflects the severity of the disease. A measure of replication capacity may be obtained by monitoring viral replication in an ideal environment, with many cellular targets, no exogenous or endogenous inhibitors, and no immune system responses against the virus (Barbour et al., 2002; Segal et al., 2004).

Genotypes of interest correspond to codons in the protease and reverse transcriptase regions of the viral strand. The protease (PR) enzyme affects the reproductive cycle of the virus by breaking protein peptide bonds during viral replication. The reverse transcriptase (RT) enzyme synthesizes double-stranded DNA from the virus' single-stranded RNA genome, thereby facilitating integration into the host's chromosome. Since the PR and RT regions are essential to viral replication, many antiretrovirals (protease inhibitors and reverse transcriptase inhibitors) have been developed to target these specific genomic locations. Studying PR and RT genotypic variation involves sequencing the corresponding HIV-1 genome regions and determining the amino acids encoded by each codon (i.e., each nucleotide triplet).

4.1.2 Description of Segal et al. (2004) HIV-1 dataset

The HIV-1 sequence dataset consists of $n = 317$ records, linking viral replication capacity (RC) with protease (PR) and reverse transcriptase (RT) sequence data, from individuals participating in studies at the San Francisco General Hospital and Gladstone Institute of Virology (Segal et al., 2004). Protease codon positions 4 to 99 (i.e., *pr4 - pr99*) and reverse transcriptase codon positions 38 to 223 (i.e., *rt38 - rt223*) of the viral strand are studied in this analysis.

The outcome/phenotype of interest is the natural logarithm of a continuous measure of replication capacity, ranging from 0.261 to 151. The M covariates correspond to the $M = 96 + 186 = 282$ codon positions in the PR and RT regions, with the number of possible codons ranging from one to ten at any given location. A majority of patients typically exhibit

one codon at each position. Codons are therefore recoded as binary covariates, with value of **zero** (or “wild-type”) corresponding to the most common codon among the $n = 317$ patients and value of **one** (or “mutation”) for all other codons. Previous biological research was used to confirm mutations and hence provide accurate PR and RT codon genotypes for each patient (hivdb.stanford.edu/cgi-bin/RTMut.cgi). The data for each of the $n = 317$ patients therefore consist of a replication capacity outcome/phenotype Y and an M -dimensional covariate vector $X = (X(m) : m = 1, \dots, M)$ of binary codon genotypes in the PR and RT HIV-1 regions.

4.2 Multiple testing procedures

The following three MTPs are applied to the HIV-1 dataset, with the aim of identifying protease and reverse transcriptase codons significantly associated with viral replication capacity: FWER-controlling single-step maxT Procedure 2, gFWER-controlling augmentation Procedure 6, and TPPFP-controlling augmentation Procedure 7.

Specifically, we wish to test for each of the $M = 282$ codon positions whether viral replication capacity Y is associated with the corresponding binary codon genotype, $X(m) \in \{0, 1\}$, $m = 1, \dots, M$. For the m th codon (i.e., m th hypothesis), the parameter of interest is the difference $\psi(m)$ in mean replication capacity of viruses with mutant and wild-type codons, that is,

$$\psi(m) \equiv E[Y|X(m) = 1] - E[Y|X(m) = 0]. \quad (23)$$

We consider two-sided tests of the null hypotheses $H_0(m) = I(\psi(m) = 0)$ of no mean difference in RC vs. the alternative hypotheses $H_1(m) = I(\psi(m) \neq 0)$ of different mean RC, based on pooled-variance two-sample t -statistics,

$$\begin{aligned} T_n(m) &\equiv \frac{\bar{Y}_1(m) - \bar{Y}_0(m) - 0}{s_p(m) \sqrt{\frac{1}{n_0(m)} + \frac{1}{n_1(m)}}}, \\ s_p^2(m) &\equiv \frac{(n_0(m) - 1)s_0^2(m) + (n_1(m) - 1)s_1^2(m)}{n_0(m) + n_1(m) - 2}, \end{aligned} \quad (24)$$

where $n_k(m)$, $\bar{Y}_k(m)$, and $s_k^2(m)$ denote, respectively, the sample sizes, sample means, and sample variances for the RC of patients with codon genotype $X(m) = k \in \{0, 1\}$ at position m . The pooled variance estimator is denoted by $s_p^2(m)$. The null hypotheses are rejected, i.e., the corresponding codons

are declared significantly associated with RC, for large absolute values of the test statistics $T_n(m)$. Note that the above two-sample t -statistics correspond to t -statistics for the univariate linear regression of the outcome Y on the binary covariates $X(m)$ (cf. one-way ANOVA).

Procedure 1 is applied to obtain a bootstrap estimate Q_{0n} of the test statistics null distribution Q_0 (with $B = 7,500$ bootstrap iterations). This estimated null distribution is used to compute adjusted p -values for the FWER-controlling single-step maxT MTP (Equation (13), Procedure 2). These FWER-controlling p -values are then used to obtain adjusted p -values for augmentation procedures controlling gFWER ($k = 5$) and TPPFP ($q = 0.1$) (for gFWER control, Equation (20), Procedure 6, and for TPPFP control, Equation (22), Procedure 7).

4.3 Software implementation in SAS

The multiple testing procedures implemented in the SAS macros of Section 3 can be applied to the HIV-1 (or a similar) dataset as follows. The parameter values specified below are those for the particular analysis of the HIV-1 dataset reported in Section 4.4 (see code in the Appendix).

1. Read in the data in the form of a SAS dataset $[y:X]$, with first column corresponding to an outcome Y (i.e., natural logarithm of viral replication capacity measure) and M subsequent columns to an M -dimensional covariate vector $X = (X(m) : m = 1, \dots, M)$ (i.e., binary codon genotypes).
2. Define the following parameters:
 - the number of rows (i.e., patients) for the dataset $[y:X]$, $n = 317$ (macro variable `&row`);
 - the number of columns, $M + 1 = 283$ (macro variable `&col`);
 - the number of bootstrap iterations for estimating the test statistics null distribution, $B = 7,500$ (macro variable `&boots`);
 - the allowed number of false positives for the gFWER-controlling AMTP, $k = 5$ (macro variable `&k`);
 - the allowed proportion of false positives for the TPPFP-controlling AMTP, $q = 0.1$ (macro variable `&q`);

- the number of hypotheses, $M = 282$ (macro variable `&nt`).
3. Apply the `%lmt` macro to compute $M = 282$ codon-specific t -statistics, T_n .
 4. Apply the `%boot` and `%bootnull` macros to create a bootstrap estimate Q_{0n} of the test statistics null distribution.
 5. Compute adjusted p -values for FWER-controlling single-step maxT Procedure 2 using the `%ssmaxT` macro.
 6. Given the adjusted p -values for the initial FWER-controlling MTP, apply the `%gfwcr` and `%tppfcr` macros to obtain adjusted p -values for augmentation procedures controlling gFWER ($k = 5$) and TPPFCR ($q = 0.1$).

Note that the Type I error rate parameters k and q and the number of bootstrap iterations B are user-supplied, in contrast to the other macro variables which are pre-determined by the dataset under consideration. For the HIV-1 dataset, the values $k = 5$, $q = 0.1$, and $B = 7,500$ were selected somewhat arbitrarily for illustration purposes.

4.4 Results

Table 3 displays the smallest 13 sorted adjusted p -values for MTPs controlling the FWER, gFWER ($k = 5$), and TPPFCR ($q = 0.1$). The total running time for the analysis was approximately 34.5 hours (Dell GX270, Intel Pentium 4, 2.8GHz, 512MB RAM).

The multiple testing procedures identified several codon positions as significantly associated with viral replication capacity. In particular, protease positions *pr32*, *pr34*, *pr43*, *pr46*, *pr47*, *pr54*, *pr55*, *pr82*, and *pr90*, and reverse transcriptase positions *rt41*, *rt184*, and *rt215*, have been singled out in previous research as related to replication capacity and/or antiretroviral resistance (Birkner et al., 2004; Segal et al., 2004; Shafer et al., 2001). It is interesting to note that the 13 codon positions with the smallest adjusted p -values in Table 3 all have negative t -statistics, suggesting that mutated codons (recoded as **one**) are associated with decreased viral replication capacity. Indeed, while mutations may allow the virus to become resistant to antiretroviral therapies, this gain may come at the cost of a decrease in its

replication capacity. The specific mutations observed in our dataset parallel those found in the literature. For example, *Vpr32I*, *Mpr46I*, *Ipr54V/L/T*, *Vpr82A/T/F/S*, and *Lpr90M*, correspond to protease positions in which mutations increase the resistance to various protease inhibitors. Here, the nomenclature *Vpr32I* refers to protease position *pr32*, with wild-type amino acid *V* (Valine) and mutated amino acid *I* (Isoleucine). Mutations in several of the identified codons also have an impact on the replication capacity of the virus. Reverse transcriptase mutation at position *rt41* (*Mrt41L*) increases azidothymidine (AZT) resistance when present with *Trt215Y/F*, i.e., with a mutated *Y* or *F* amino acid at position *rt215*. In addition, mutation *Mrt184V/I* suppresses the wild-type activity of *Trt215Y*, thus decreasing AZT resistance (Shafer et al., 2001). AZT, also known as Zidovudine, is a nucleoside reverse transcriptase inhibitor. It affects HIV's ability to replicate by producing faulty reverse transcriptase and hence inhibiting the transcription of RNA to DNA.

As illustrated in the above HIV-1 data analysis, the multiple testing methodology of Section 2 provides simple and flexible procedures for identifying specific codons, or areas on the viral strand, that are significantly associated with replication capacity. Our results are consistent with previous research and other analyses of this HIV-1 dataset. The reader is referred to earlier articles by Birkner et al. (2004) and Segal et al. (2004) for alternative statistical analyses and biological discussion of a related HIV-1 dataset. Recall that a crude binary (wild-type vs. mutated) coding was used for codon genotypes; a more sensitive analysis may be achieved by using individual amino acids or grouping amino acids based on their biochemical properties.

Other multiple testing software

A number of bootstrap- and permutation-based FWER-controlling MTPs are implemented in PROC MULTTEST (SAS OnlineDoc, Version 9, [v9doc.sas.com/sasdoc](http://www.sas.com/sasdoc)).

The resampling-based multiple testing procedures discussed in this and related articles (Dudoit and van der Laan, 2005; Dudoit et al., 2004a,b; van der Laan et al., 2004a,b; Pollard and van der Laan, 2004) are implemented in the open-source R package `multtest`, released as part of the Bioconductor Project (Pollard et al. (2005); `multtest` package Version 1.5.2, Bioconductor Release 1.5, www.bioconductor.org).

Acknowledgments

We are most grateful to Mark R. Segal (Department of Epidemiology and Biostatistics, University of California, San Francisco) for providing access to the HIV-1 dataset and for helpful discussions on its analysis. We would also like to thank David Lein (Division of Biostatistics, University of California, Berkeley) and Phil Spector (Department of Statistics, University of California, Berkeley) for their valuable advice on SAS programming.

References

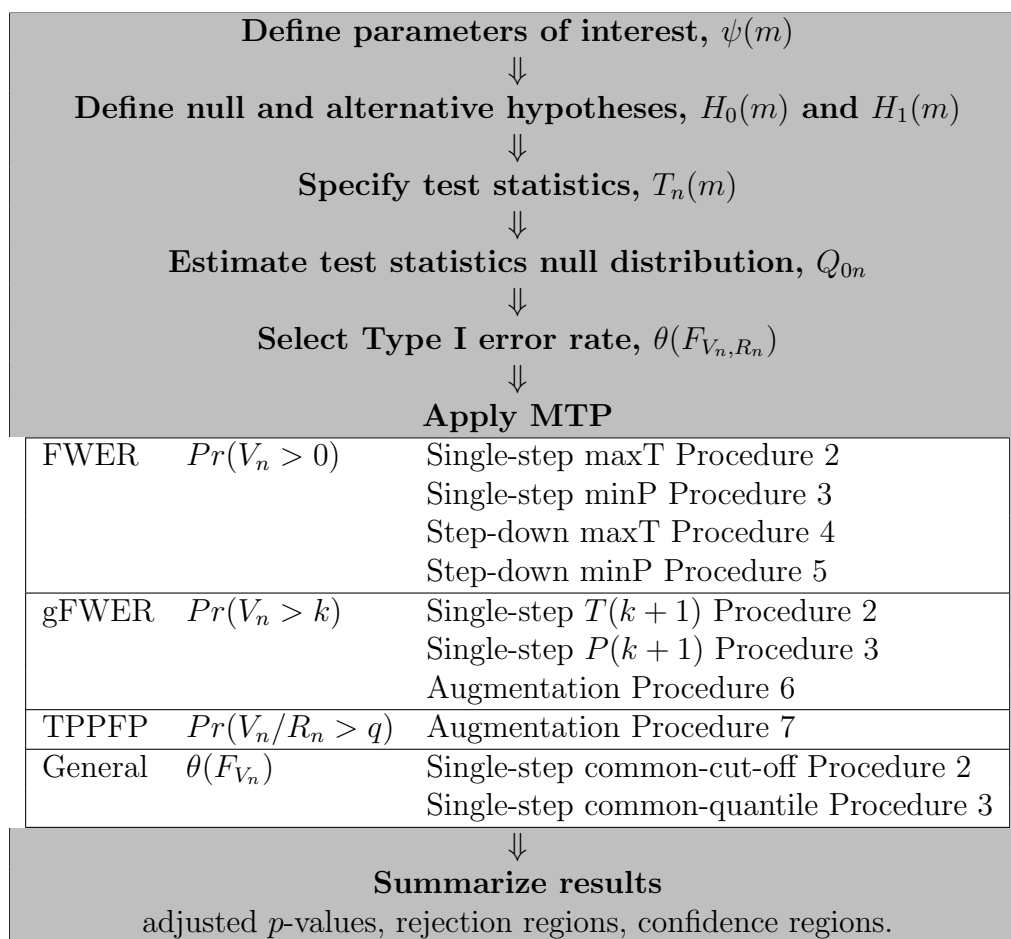
- J. D. Barbour, T. Wrin, R. M. Grant, J. N. Martin, M. R. Segal, C. J. Petropoulos, and S. G. Deeks. Evolution of Phenotypic Drug Susceptibility and Viral Replication Capacity during Long-Term Virologic Failure of Protease Inhibitor Therapy in Human Immunodeficiency Virus-Infected Adults. *Journal of Virology*, 76(21):11104–11112, 2002.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300, 1995.
- M. D. Birkner, S. E. Sinisi, and M. J. van der Laan. Multiple testing and data adaptive regression: An application to HIV-1 sequence data. Technical Report 161, Division of Biostatistics, University of California, Berkeley, 2004. URL www.bepress.com/ucbbiostat/paper161.
- S. Dudoit and M. J. van der Laan. *Multiple Testing Procedures and Applications to Genomics*. Springer, 2005. (In preparation).
- S. Dudoit, M. J. van der Laan, and M. D. Birkner. Multiple testing procedures for controlling tail probability error rates. Technical Report 166, Division of Biostatistics, University of California, Berkeley, 2004a. URL www.bepress.com/ucbbiostat/paper166.
- S. Dudoit, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 13, 2004b. URL www.bepress.com/sagmb/vol3/iss1/art13.

- S. Keleş, M. J. van der Laan, S. Dudoit, and S. E. Cawley. Multiple testing methods for ChIP-Chip high density oligonucleotide array data. Technical Report 147, Division of Biostatistics, University of California, Berkeley, 2004. URL www.bepress.com/ucbbiostat/paper147.
- K. S. Pollard, S. Dudoit, and M. J. van der Laan. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, chapter Multiple Testing Procedures and Applications to Genomics. Springer-Verlag, New York, 2005. URL www.bepress.com/ucbbiostat/paper164. (Submitted).
- K. S. Pollard and M. J. van der Laan. Choice of a null distribution in resampling-based multiple testing. *Journal of Statistical Planning and Inference*, 125(1–2):85–100, 2004.
- M. R. Segal, J. D. Barbour, and R. M. Grant. Relating HIV-1 Sequence Variation to Replication Capacity via Trees and Forests. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 2, 2004. URL www.bepress.com/sagmb/vol3/iss1/art2.
- R. W. Shafer, K. M. Dupnik, M. A. Winters, and S. H. Eshleman. A Guide to HIV-1 Reverse Transcriptase and Protease Sequencing for Drug Resistance Studies. In *HIV Sequencing Compendium*, pages 83–133. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2001.
- M. J. van der Laan, S. Dudoit, and K. S. Pollard. Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 15, 2004a. URL www.bepress.com/sagmb/vol3/iss1/art15.
- M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 14, 2004b. URL www.bepress.com/sagmb/vol3/iss1/art14.
- P. H. Westfall and S. S. Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons, 1993.

Table 1: *Type I and Type II errors in multiple hypothesis testing.*

		Null hypotheses		
		not rejected	rejected	
Null hypotheses	true	$ \mathcal{R}_n^c \cap \mathcal{H}_0 $	$V_n = \mathcal{R}_n \cap \mathcal{H}_0 $ (Type I errors)	$h_0 = \mathcal{H}_0 $
	false	$U_n = \mathcal{R}_n^c \cap \mathcal{H}_1 $ (Type II errors)	$ \mathcal{R}_n \cap \mathcal{H}_1 $	$h_1 = \mathcal{H}_1 $
		$M - R_n$	$R_n = \mathcal{R}_n $	M

Table 2: *Multiple hypothesis testing flowchart.*



COBRA
A BEPRESS REPOSITORY

Collection of Biostatistics
Research Archive

Table 3: *HIV-1 dataset*. Sorted adjusted p -values and t -statistics for multiple testing procedures controlling the FWER, gFWER ($k = 5$), and TPPFP ($q = 0.1$).

Codons	t -statistics	Adjusted p -values		
		FWER SS maxT	gFWER ($k = 5$) AMTP	TPPFP ($q = 0.1$) AMTP
<i>pr32</i>	-9.755	0.0001	0	0.0001
<i>pr47</i>	-9.579	0.00133	0	0.00133
<i>pr34</i>	-8.843	0.00867	0	0.00867
<i>pr55</i>	-8.15	0.0104	0	0.0104
<i>pr90</i>	-6.237	0.0396	0	0.0396
<i>rt184</i>	-6.162	0.0431	0.0001	0.0431
<i>pr43</i>	-6.118	0.0444	0.00133	0.0444
<i>pr54</i>	-5.539	0.078	0.00867	0.078
<i>rt41</i>	-5.225	0.978	0.0104	0.978
<i>pr46</i>	-5.224	0.098	0.0396	0.098
<i>pr82</i>	-4.521	0.1678	0.0431	0.0978
<i>rt215</i>	-4.479	0.174	0.0444	0.1678
<i>rt121</i>	-4.07	0.238	0.078	0.174



COBRA
A BEPRESS REPOSITORY

Collection of Biostatistics
Research Archive

Appendix: SAS code

```

/*****
/* Read in HIV-1 SAS dataset, [y:X] */

options nonotes;
%let row = 317;      /* number of patients */
%let col = 283;     /* first column corresponds to RC outcome, remaining columns to PR and
%let boots = 7500; /* number of bootstrap samples */
%let k = 5;         /* number of allowed false positives for gFWER AMTP */
%let q = 0.05;     /* proportion of allowed false positives for TPPFP AMTP */
%let nt= 282;      /* number of tested hypotheses */

libname resample "c:\hivexample";

/*****
/* lmt: t-statistics for univariate linear regression model */

%macro lmt;
  %do a = 2 %to &col;
    proc reg data=resample.hivdata noprint outest=outest&a tableout;
      model VAR1=VAR&a;
      data outest&a(rename=(VAR&a=t)); set outest&a; where _type_='T'; keep VAR&a;
      proc append base=tstats data=outest&a;
    %end;
  %mend;
%lmt;
proc print data=tstats; title "t-statistics for data";
run;

/*****
/* boot and bootnull: Bootstrap test statistics null distribution */

%macro boot;
%do j=1 %to &boots;
proc surveysselect noprint data=resample.hivdata out=datanew
  seed=&j
  method=urs
  rep=&row

```

```

        samsize=1
        stats;
%do a = 2 %to &col;
proc reg data=datanew noprint outest=outest&a tableout;
        model VAR1=VAR&a;
data outest&a(rename=(VAR&a=t)); set outest&a; where _type_='T'; keep VAR&a;
proc append base=tstatsB data=outest&a;
%end;
%end;
%mend;
%boot;
quit;

/*****/
%macro bootnull;
proc iml;
        use tstatsB;
        read all var {t} into x;
        close tstatsB;
        nt=&nt;
        bt=&boots;
        tB=shape(x,bt,nt);
        b=tB[:,];
        bb=J(bt,nt,0);
        do i=1 to bt;
                bb[i,]=(tB[i,] - b);
        end;
        bb2=shape(bb,bt*nt,1);
        create Qo from bb2;
        append from bb2;
        close Qo;
        quit;
%mend;
%bootnull;

/*****/
/* ssmaxT: Single-step maxT procedure */

%macro ssmaxT;
proc iml;

```

```

use Qo;
read all into bb;
close Qo;
nt=&nt;
bt=&boots;
Qo=shape(bb,bt,nt);
mx=J(1,bt,0);
do i=1 to bt;
    mb=abs(Qo[i,]);
    mx[,i]=max(mb);
end;
c=mx;
brank=rank(mx);
mx[brank]=c;
/*adjusted p-values*/
use tstats;
read all var {t} into t;
close tstats;
print t;
pval=J(1,nt,0);
do j=1 to nt;
    tmp=J(1,bt,0);
    do i=1 to bt;
        if abs(t[j]) < mx[i] then tmp[i]=1;
    end;
    st=sum(tmp);
    pval[j]=st/bt;
end;
create fwer from pval;
append from pval;
close fwer;
quit;
proc print data=fwer;
run;
%mend;
%ssmaxT;

/*****
/* gfwer: gFWER-controlling augmentation procedure */

```

```

%macro gfwer;
proc iml;
    use fwer;
    read all into pval;
    close fwer;
    k=&k;
    nt=&nt;
    gp=J(1,nt,0);
    j=k+1;
    c=pval;
    brank=rank(pval);
    pval[brank]=c;
    do i=1 to (nt-k);
        gp[j]=pval[i];
        j=j+1;
    end;
    gg=gp[brank];
    gpval=shape(gg,1,nt);
    create gfwer from gpval;
    append from gpval;
    close gfwer;
    quit;
proc print data=gfwer;
run;
%mend;
%gfwer;

/*****/
/* tppfp: TPPFP-controlling augmentation procedure */

%macro tppfp;
proc iml;
    use fwer;
    read all into pval;
    close fwer;
    q=&q;
    nt=&nt;
    c=pval;
    brank=rank(pval);
    pval[brank]=c;

```

```

tp=J(1,nt,0);
do i=1 to nt;
    m=ceil(i*(1-q));
    tp[i]=pval[m];
end;
tt=tp[brank];
tpval=shape(tt,1,nt);
create tppfp from tpval;
append from tpval;
close tppfp;
quit;
proc print data=tppfp;
run;
%mend;
%tppfp;

/*****/

```

