Comment on “A note on the adaptive control of false discovery rates” by M.A. Black

John D. Storey
Department of Biostatistics, University of Washington, Seattle, USA

Abstract. In a recent article, Black (2004) makes comparisons between two approaches to false discovery rates that have recently been proposed in the literature. I show that his conclusions do not necessarily hold in general, and that they are not as straightforward as implied. I also show that the connections that he makes between the two approaches have already been shown in a number of previous publications, including one of the original publications that he discusses (Storey 2002). Black’s (2004) heuristic criticisms and observations are clarified in the context of previously published theoretical and methodological developments.

Address for correspondence: John D. Storey, Department of Biostatistics, University of Washington, Seattle, WA 98195, USA. E-mail: jstorey@u.washington.edu

In a recent article, Black (2004) makes several comparisons between two approaches to false discovery rates that have recently been proposed in the literature. Benjamini & Hochberg (1995) proposed the first false discovery rate controlling procedure, which was introduced by Simes (1986) in the context of the family-wise error rate. In their approach, they fix an acceptable false discovery rate level, and perform a procedure to estimate a corresponding p-value threshold. When the null p-values are independent, it is shown that this thresholding procedure yields an expected false discovery rate that is less than or equal to the pre-chosen level, regardless of the number of true null hypotheses. Because of this property, the procedure is said to provide “strong control” of the false discovery rate. Subsequently, Benjamini & Yekutieli (2001) have shown that this procedure continues to provide strong control under “positive dependence” of the null p-values. It is immediately clear from Benjamini & Hochberg (1995) that their procedure can be made less conservative by incorporating an estimate of the number of true null hypotheses into the procedure. Benjamini & Hochberg (2000) propose such a procedure; however, this procedure is heuristically motivated and no theoretical justification has yet been provided.

In Storey (2002), I take a different approach to the false discovery rate problem. Instead of fixing the error rate and estimating an appropriate threshold, I fix the threshold and develop a point estimate of the false discovery rate. In that article, I claim that this provides an alternative approach to studying and estimating false discovery rates. I also show in Storey (2002) that the proposed procedure is closely related to the procedure proposed in Benjamini & Hochberg (1995). Section 4 of Storey (2002) makes an explicit connection; on page 485, I state “The operational difference between [my estimate] and the Benjamini & Hochberg (1995) procedure is the inclusion of [an estimate of the proportion of null hypotheses].” I show that my estimate is algebraically equivalent to the Benjamini & Hochberg (1995) procedure except for the inclusion of the estimate of the proportion of true null hypotheses. I also state that Benjamini & Hochberg (2000) have taken a similar approach. Efron & Tibshirani (2002), Genovese & Wasserman (2002), and Storey et al. (2004) are subsequent articles that show this connection in detail as well. In Storey et al. (2004) we show that a slightly modified version of the estimate proposed in Storey (2002) provides strong control of the false discovery rate, in the Benjamini & Hochberg (1995) sense. In the abstract we say, “In this work, we show in both finite sample and asymptotic settings that the goals of the two approaches are essentially equivalent.”

Black (2004) claims that he has discovered this connection. The abstract states, “In this work it is demonstrated that, if the original step-up procedure of Benjamini and Hochberg is modified to exercise adaptive control of the false discovery rate, its performance is virtually identical to that of the fixed rejection region approach.” As stated above, this was clearly stated in the original paper (Storey 2002) proposing the

---

1This paper originally appeared as UC-Berkeley Statistics Technical Report 623 in July 2002.
fixed rejection region approach. Black (2004) seems to have missed the major point of Storey (2002), that
taking the arguably more straightforward fixed rejection region approach leads to similar procedures.

Black (2004) also claims that the procedures proposed in Benjamini & Hochberg (2000) and Storey (2002)
are virtually the same. Specifically, he says, “the fixed rejection region method (Storey, 2002) and adaptive
FDR control (Benjamini and Hochberg, 2000) essentially provide the same level of control of the error rate,
despite the fact that they approach the problem from opposite directions. Given that these techniques
appear radically different at first glance, their similarity should provide some comfort for researchers trying
to decide which method is more appropriate.” From his simple simulation, this appears to be the case.
However, note that strong control has been shown (Storey et al. 2004) for the estimate in Storey (2002),
but strong control has not been shown for the procedure in Benjamini & Hochberg (2000). Therefore,
it is not safe to conclude that they essentially provide the same level of false discovery rate control in
general. Also, as stated above, it has already been shown (Benjamini & Hochberg 2000, Storey 2002, Efron
& Tibshirani 2002, Genovese & Wasserman 2002, Storey et al. 2004) that these procedures, and any other
procedure incorporating information about the number of true null hypotheses, are related. This is not the
same as providing strong control, and one has to be careful about obfuscating these two points.

The comparison between Benjamini & Hochberg (2000) and Storey (2002) is not as straightforward as
adaptive because the form of the estimate for the number of true null hypotheses depends on the data.
However, Black (2004) fixes the tuning parameter in Storey’s (2002) estimate to \(\lambda = 0.5\), which puts this
estimate at a disadvantage. Strong control has been shown to hold in the finite sample case for Storey’s
(2002) procedure for a fixed tuning parameter. Again, no strong control has been shown to hold for the
procedure for choosing the tuning parameter in Storey’s (2002) estimate that leads to underestimation. On
the contrary, we have shown in Storey et al. (2004) that the procedure continues to provide strong control
in the asymptotic setting as long as only a finite number of tuning parameter values are considered. In the
finite sample setting, it is always possible to find a numerical example where a procedure such as this goes
wrong, so Black (2004) has not provided us with any information that holds in general.

In conclusion, Black (2004) claims to have made new connections between the two approaches to false
discovery rates. These connections have been spelled out in detail in several previous publications (Storey
existing procedures are equivalent (Benjamini & Hochberg 2000, Storey 2002). This is also not the case
because one of them has been shown to provide strong control of the false discovery rate, and the other
has not. Moreover, Benjamini & Hochberg (2000) is a data-adaptive procedure and Storey (2002), as it is
presented in Black’s (2004) simulations, is not. Black (2004) does not consider the new results about Storey’s
(2002) procedure that have shown in Storey et al. (2004). These new finite sample and large sample results
provide a rigorous justification for the method under a variety of conditions, and these results have not
been verified for other procedures that heuristically appear to be similar. Black’s (2004) claims, which are
implicitly stated as generally holding true, are justified by intuition-based and simulation-based comparisons,
rather than through rigorous theoretical development.

References


\(^2\)It should be noted that Benjamini, Krieger, & Yekutieli in an unpublished manuscript have developed a procedure that
incorporates an estimate of the number of true null hypotheses, and has also been proven to provide strong control of the false
discovery rate (personal communication).


