Biostatistics (2013), **0**, 0, *pp.* 1–16
doi:10.1093/biostatistics/SWFDR˙Supp

Empirical estimates suggest most published research is true: Supplementary Materials

LEAH R. JAGER[1] and JEFFREY T. LEEK[2*]

1. Department of Mathematics, United States Naval Academy, Annapolis, MD 21402, USA, and

2. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

jleek@jhsph.edu

1. The two-groups model and the false discovery rate

A common model for p-values from multiple hypothesis tests is the "two-groups" model (Efron and Tibshirani, 2002), where the p-values are assumed to be drawn from the mixture distribution:

$$p \sim \pi_0 f_0 + (1 - \pi_0) f_1$$

where π_0 is the proportion of p-values corresponding to tests where the null hypothesis is true, $f_0(\cdot)$ is the density of the p-values under the null hypotheses and $f_1(\cdot)$ is the density of p-values under the alternative hypothesis. In general, each test may have a corresponding null and alternative distribution, in which case both f_0 and f_1 may be represented by mixture distributions.

This model is frequently used in genomics (Newton *et al.*, 2004), or brain imaging studies (Genovese, Lazar, and Nichols, 2002), where thousands of variables are measured simultaneously with the goal of estimating the statistical significance of a group of variables (Storey and Tibshirani, 2003). This model simultaneously estimates π_0, the fraction of results that are drawn from

*To whom correspondence should be addressed.

the null distribution. Alternatively, $1-\pi_0$ represents the fraction of p-values from the alternative distribution.

1.1 *A modified two-groups model for p-values < 0.05*

It is assumed that p-values are distributed $\mathcal{U}(0,1)$ under the null hypothesis (Lehmann, 1991), so $f_0(p) = \mathbb{1}[p \in (0,1)]$. A common choice for parameterizing the alternative distribution is the Beta distribution $\beta(a,b)$, with $f_1(p) = \beta(p;a,b)$ (Allison *et al.*, 2002). But the reported p-values in the published literature do not represent the complete distribution of p-values from all hypotheses tested. Indeed, most p-values reported in abstracts are less than 0.05, the conventional threshold for statistical significance used by most journals ($p \leqslant 0.05$ for 81% of the p-values in our data). So we define a new two-groups model for p-values reported in the scientific literature conditional on the event $\{p \leqslant 0.05\}$:

$$p \sim \pi_0^r f_0^r + (1-\pi_0^r) f_1^r \tag{1.1}$$

here π_0^r is the proportion of p-values corresponding to reported null hypotheses, $f_0^r(\cdot)$ is the distribution of reported p-values under the null hypothesis, conditional on $\{p \leqslant 0.05\}$, and $f_1^r(\cdot)$ is defined analogously for the alternative.

If $p \sim \mathcal{U}(0,1)$ under the null hypothesis, then $p|\{p \leqslant 0.05\} \sim \mathcal{U}(0, 0.05)$, and if $p \sim \beta(a,b)$ under the alternative hypothesis then: $p|\{p \leqslant 0.05\} \sim \frac{1}{F_{a,b}(0.05)}\beta(a,b) = t\beta(a,b)$, both by properties of conditional distributions (Casella and Berger, 2002).

We can fit the modified two groups model (1.1) using the EM-algorithm (Dempster, Laird, and Rubin, 1977). First, we introduce the (unobserved) indicator z_i which is equal to 1 if the p-value corresponds to a null hypothesis and 0 if the p-value corresponds to the alternative hypothesis. We can then write the complete data likelihood (for m total p-values) as:

$$L(a,b,\pi_0|p_i,z_i) = \prod_{i=1}^{m} [\pi_0 \cdot 20]^{z_i} \left[(1-\pi_0)\frac{1}{F_{a,b}(0.05)}\beta(p_i;a,b)\right]^{1-z_i} \tag{1.2}$$

where the indicator for the null distribution has been suppressed for ease of notation. This is a standard two-groups mixture model and can be fit with the EM-algorithm directly.

1.2 *Rounding and truncation of p-values*

There are two key issues that prevent direct use of the likelihood (1.2). The first is that reported p-values are frequently left-censored, so if the calculated p-value is $p = 1.134 \times 10^{-12}$ then the reported p-value is often $p < 0.001$. We treat these observations as left-censored and introduce the indicator function δ_i, which is equal to 1 if p-value i is censored and 0 otherwise. Since censored p-values are reported with $<$ or $\leqslant$ signs in the literature, the truncation indicator is directly observed. Following the usual approach for modeling parametric survival data (Kleinbaum and Klein, 2005) (with time running backward) we can modify equation (1.2) to deal with the contribution of the censored p-values:

$$L(a, b, \pi_0 | p_i, z_i, \delta_i) = \prod_{i=1}^{m} \left[\pi_0 \left(20\right)^{1-\delta_i} \left(\int_0^{p_i} 20 \; dt \right)^{\delta_i} \right]^{z_i}$$
$$\cdot \left[(1-\pi_0) \left(\frac{1}{F_{a,b}(0.05)} \beta(p_i; a, b) \right)^{1-\delta_i} \left(\int_0^{p_i} \frac{1}{F_{a,b}(0.05)} \beta(t; a, b) dt \right)^{\delta_i} \right]^{1-z_i}$$

The second issue is that reported p-values are frequently rounded to two significant digits, so if the calculated p-value is $p = 0.033333$, the reported p-value is often $p = 0.03$. Histograms of the mined p-values show clear spikes at the values 0.01, 0.02, 0.03, 0.04, and 0.05 (Figure 3). Any p-value that is reported exactly as a value with only two significant digits was assumed to be rounded.

We introduce a second indicator, r_i, which is equal to 1 if the p-value is rounded and 0 otherwise. We consider the rounding indicator, like the censored indicator, to be observed. We then modify the likelihood to include a multinomial term for the rounded p-values with values at 0, 0.01, 0.02, 0.03, 0.04 and 0.05. The probability of each value under the null and alternative

hypothesis are calculated by integrating the null (or alternative) distribution over the interval of values that round to the two digit number (MacDonald and Pitcher, 1979; Wengrzik and Timm, 2011). For example, under the null, the probability of a reported value of 0.01 is: $\int_{0.005}^{0.015} f_0(x)dx$. Using this framework we can define the probabilities of each rounded value under the null and alternative hypotheses:

$$q_j^k = \int_{l(k)}^{u(k)} f_j^r(x)dx$$

where $k = 0, 1, 2, 3, 4, 5$, the upper and lower bounds are defined by the rounding regions, and $j = 0, 1$ for the null and alternative hypotheses, respectively. Since these q_j^k are conditional on whether the observed p-values are null or alternative, the unconditional probabilities for this multinomial distribution are $\pi_j q_j^k$, where $\pi_1 = 1 - \pi_0$.

For each of the k values, we observe the number of p-values rounded to that value, which we denote $\vec{n} = (n^0, n^1, n^2, n^3, n^4, n^5)$ and we also introduce a second, unobserved, variable equal to the number of observations from the null distribution corresponding to each value, which we denote $\vec{n}_0 = (n_0^0, n_0^1, n_0^2, n_0^3, n_0^4, n_0^5)$. The corresponding values for the alternative distribution are then $\vec{n}_1 = (n^0 - n_0^0, n^1 - n_0^1, n^2 - n_0^2, n^3 - n_0^3, n^4 - n_0^4, n^5 - n_0^5)$. With the addition of these variables, the complete data likelihood is:

$$\begin{aligned} L(a, b, \pi_0 | p_i, z_i, \delta_i, r_i) = & \left(\prod_{\{i:r_i=0\}} \left[\pi_0 \left(20\right)^{1-\delta_i} \left(\int_0^{p_i} 20 \; dt \right)^{\delta_i} \right]^{z_i} \right. \\ & \cdot \left. \left[(1-\pi_0) \left(\frac{1}{F_{a,b}(0.05)} \beta(p_i; a, b) \right)^{1-\delta_i} \left(\int_0^{p_i} \frac{1}{F_{a,b}(0.05)} \beta(t; a, b) dt \right)^{\delta_i} \right]^{1-z_i} \right) \\ & \cdot M(\vec{n}_0, \pi_0 \vec{q}_0) \cdot M(\vec{n}_1, (1-\pi_0)\vec{q}_1) \end{aligned} \tag{1.3}$$

1.3 *An EM Algorithm for fitting the two-groups model*

To estimate the parameters in model (1.3) we apply the EM algorithm. The first step is to calculate the expected value of the complete data log-likelihood conditional on the observed data. The log-likelihood is:

$$\begin{aligned}\ell(a,b,\pi_0) = \sum_{\{i:r_i=0\}} & [z_i(1-\delta_i)\log(20) + z_i\log(\pi_0) \\ & + z_i\delta_i \log\left(\int_0^{p_i} 20\ dt\right) + (1-z_i)\log(1-\pi_0) \\ & + (1-z_i)(1-\delta_i)\log\left(\frac{1}{F_{a,b}(0.05)}\beta(p_i;a,b)\right) \\ & + (1-z_i)\delta_i \log\left(\int_0^{p_i} \frac{1}{F_{a,b}(0.05)}\beta(t;a,b)dt\right)\Big] \\ & + \sum_{k=0}^{5} n_0^k \log(\pi_0) + n_0^k \log(q_0^k) \\ & + \sum_{k=0}^{5} (n^k - n_0^k)\log(1-\pi_0) + (n^k - n_0^k)\log(q_1^k)\end{aligned}$$

1.3.1 *E-step* The log-likelihood is linear in the unobserved variables z_i and $\vec{n}_0$, so to calculate the expectation of the complete-data log-likelihood we calculate the expectation of these variables conditional on the observed data.

- If the p-value is not censored or rounded ($\delta_i = 0$, $r_i = 0$) the expectation is:

$$\begin{aligned}\mathrm{E}[z_i|p_i] &= \Pr(z_i = 1|p_i) \\ &= \frac{\Pr(p_i|z_i=1)\Pr(z_i=1)}{\Pr(p_i|z_i=1)\Pr(z_i=1) + \Pr(p_i|z_i=0)\Pr(z_i=0)} \\ &= \frac{20\pi_0}{20\pi_0 + (1-\pi_0)\left(TruncBeta(p_i|a,b)\right)}\end{aligned}$$

- If the p-value is censored but not rounded ($\delta_i = 1$, $r_i = 0$) the expectation is:

$$\begin{aligned}
\mathrm{E}[z_i | P_i < p_i] &= \Pr(z_i = 1 | P_i < p_i) \\
&= \frac{\Pr(P_i < p_i | z_i = 1)\Pr(z_i = 1)}{\Pr(P_i < p_i | z_i = 1)\Pr(z_i = 1) + \Pr(P_i < p_i | z_i = 0)\Pr(z_i = 0)} \\
&= \frac{\pi_0 \int_0^{p_i} 20 \, dt}{\pi_0 \int_0^{p_i} 20 \, dt + (1 - \pi_0) \left(\int_0^{p_i} TruncBeta(t | a, b) dt \right)}
\end{aligned}$$

Similarly we can calculate the expectation of each element of $\vec{n}_0$ as

$$\begin{aligned}
\mathrm{E}[n_0^k | \vec{n}, \pi_0] &= n^k \times \Pr(\text{Value } k \text{ and Null}) = n^k \Pr(\text{Value } k | \text{Null}) \Pr(\text{Null}) \\
&= n^k q_0^k \pi_0.
\end{aligned}$$

1.3.2 *M-step* The log-likelihood can be maximized directly for the parameter π_0 by solving the equation:

$$\frac{\partial \ell}{\partial \pi_0} = \frac{\sum_{\{i:r_i=0\}} z_i}{\pi_0} - \frac{\sum_{\{i:r_i=0\}} 1 - z_i}{1 - \pi_0} + \frac{\sum_{k=0}^5 n_0^k}{\pi_0} - \frac{\sum_{k=0}^5 (n^k - n_0^k)}{1 - \pi_0} = 0$$

which leads to the estimate

$$\hat{\pi}_0 = \frac{\sum_{\{i:r_i=0\}} z_i + \sum_{k=0}^5 n_0^k}{\sum_i \mathbb{1}\{r_i = 0\} + \sum_{k=0}^5 n^k}$$

The integrals in $\vec{q}_j$ mean that the log-likelihood can not be directly maximized for the parameters (a, b). At each step of the EM algorithm, we numerically maximize the part of the log-likelihood equation that depends on (a, b),

$$\begin{aligned}
&\sum_{\{i:r_i=0\}} \left[(1 - z_i)(1 - \delta_i) \log \left(\frac{1}{F_{a,b}(0.05)} \beta(p_i; a, b) \right) \right. \\
&\left. + (1 - z_i)\delta_i \log \left(\int_0^{p_i} \frac{1}{F_{a,b}(0.05)} \beta(t; a, b) dt \right) \right] \\
&+ \sum_{k=0}^5 (n^k - n_0^k) \log(q_1^k),
\end{aligned}$$

with respect to (a, b) using the `mle` R function in the `stats4` package.

For the results in our paper, the E-step and M-step are iterated 100 times. We applied the EM-algorithm to calculate the false discovery rate in each journal in each year. We also applied the EM-algorithm to calculate the rate of false discovery rate in each journal considering all years and to calculate the overall false discovery rate.

2. Bootstrap standard errors for the false discovery rate

We calculated bootstrap standard errors for the estimates $\hat{\pi}_0$ by sampling with replacement sets of variables $\{(p_i, \delta_i, r_i)\}$ of size m, the number of observed p-values, and applying the EM algorithm to each bootstrap sample. The bootstrapped standard deviation is an estimate of the standard deviation of the false discovery rate estimates (Efron and Tibshirani, 1993).

3. p-value reporting and the conditional p-value model

Our analysis is focused on estimating a false discovery rate across multiple studies. In this case, not all p-values may be reported. The critical component of conservatively estimating the false discovery rate is that the null distribution is accurately modeled. The following proposition makes clear the conditions under which our conditional model is appropriate for the null distribution of the sampled p-values.

Proposition 1 *Let $p_1, \ldots, p_{m_0}$ be p-values corresponding to tested null hypotheses. Let $x_i = 1(p_i < \alpha)$ be the indicator functions that the p-values are less than threshold α. Let $y_i = 1(p_i$ is reported$)$ be the indicators that the p-values are reported. If the null p-values are correctly calculated and $y_i \perp p_i | x_i$ then the reported p-values less than α are distributed $U(0, \alpha)$.*

Proof If the null p-values are correctly calculated then $p_i \sim U(0, 1)$. The distribution of the reported p-values less than α is $f(p_i | y_i = 1, x_i = 1) = f(p_i | x_i = 1)$ by the conditional independence

assumption. Since $p_i \sim U(0,1)$, this reduces to $f(p_i|p_i < 0.05)$. $f(p_i) = 1(p_i \in (0,1))$ so by the properties of uniform random variables $f(p_i|x_i = 1) = 1(p_i \in (0,\alpha))$. □

In particular, Proposition 1 implies that the following cases result in accurate models of the null distribution using our proposed conditional model: (1) investigators report all p-values less than α as significant, (2) investigators report the first p-value they observe less than α as significant, (3) investigators report all p-values less than α as significant, only when their intuition suggests they should report that p-value, as long as that intuition is based on separate biological knowledge that the result is interesting independent of the p-value. However, there are some cases where the conditional model is not theoretically correct. For example, if investigators calculate many p-values and only report the minimum p-value.

4. Collecting p-values from reported studies

Our p-value data was collected using the code in the supplementary `R` script `getPvalues.R`. The script scrapes the Pubmed website for abstracts corresponding to papers published in *The American Journal of Epidemiology*, *The Journal of the American Medical Association*, *The New England Journal of Medicine*, *The British Medical Journal*, and *The Lancet*. These p-values were scraped from http://www.ncbi.nlm.nih.gov/pubmed/ on January 24, 2012, a small amount of manual correction was performed and is detailed in the `R` script.

Any p-values that were calculated as NA were removed from the analysis and the analysis is based on only the p-values less than 0.05 as described above.

5. Collecting data on the number of submissions to medical journals

We directly contacted the editors of *The American Journal of Epidemiology*, *The Journal of the American Medical Association*, *The New England Journal of Medicine*, *The British Medical*

Journal, and *The Lancet*, who supplied the data on the number of submitted manuscripts to their respective journals.

6. Manual validation of data for a random set of abstracts

Ten Pubmed ID's were selected using the random number generator in the accompanying R script. Here we report the collected P-values and abstracts for these Pubmed ID's. P-values appearing in abstracts are highlighted in blue.

1. **Pubmed ID: 15950715, Collected P-value(s): 0.07**
   BACKGROUND Survivors of malignant disease in childhood who have had radiotherapy to the head, neck, or upper thorax have an increased risk of subsequent primary thyroid cancer, but the magnitude of risk over the therapeutic dose range has not been well established. We aimed to quantify the long-term risk of thyroid cancer after radiotherapy and chemotherapy.

   METHODS: In a nested case-control study, 69 cases with pathologically confirmed thyroid cancer and 265 matched controls without thyroid cancer were identified from 14,054 5-year survivors of cancer during childhood from the Childhood Cancer Survivor Study cohort. Childhood cancers were diagnosed between 1970 and 1986 with cohort follow-up to 2000.

   FINDINGS: Risk of thyroid cancer increased with radiation doses up to 20-29 Gy (odds ratio 9.8 [95% CI 3.2-34.8]). At doses greater than 30 Gy, a fall in the dose-response relation was seen. Both the increased and decreased risks were more pronounced in those diagnosed with a first primary malignant disease before age 10 years than in those older than 10 years. Furthermore, the fall in risk remained when those diagnosed with Hodgkin's lymphoma were excluded. Chemotherapy for the first cancer was not associated with thyroid-cancer risk, and it did not modify the effect of radiotherapy. 29 (42%) cases had a first diagnosis of Hodgkin's lymphoma compared with 49 (19%) controls. 11 (42%) of those who had Hodgkin's lymphoma had subsequent thyroid cancers smaller than 1 cm compared with six (17%) of those who had other types of childhood cancer (p=0.07).

   INTERPRETATION: The reduction in radiation dose-response for risk of thyroid cancer after childhood exposure to thyroid doses higher than 30 Gy is consistent with a cell-killing effect. Standard long-term follow-up of patients who have had Hodgkin's lymphoma for detection of thyroid cancer should also be undertaken for survivors of any cancer during childhood who received radiotherapy to the thorax or head and neck region.

2. **Pubmed ID: 19181729, Collected P-value(s): 0.001,0.08,0.02**
   OBJECTIVE: To assess the clinical effectiveness and cost effectiveness of a policy to provide breastfeeding groups for pregnant and breastfeeding women.

   DESIGN: Cluster randomised controlled trial with prospective mixed method embedded case studies to evaluate implementation processes.

   SETTING: Primary care in Scotland.

   PARTICIPANTS: Pregnant women, breastfeeding mothers, and babies registered with 14 of 66 eligible clusters of general practices (localities) in Scotland that routinely collect breastfeeding outcome data.

   INTERVENTION: Localities set up new breastfeeding groups to provide population coverage; control localities did not change group activity.

MAIN OUTCOME MEASURES: Primary outcome: any breast feeding at 6-8 weeks from routinely collected data for two pre-trial years and two trial years. Secondary outcomes: any breast feeding at birth, 5-7 days, and 8-9 months; maternal satisfaction.

RESULTS: Between 1 February 2005 and 31 January 2007, 9747 birth records existed for intervention localities and 9111 for control localities. The number of breastfeeding groups increased from 10 to 27 in intervention localities, where 1310 women attended, and remained at 10 groups in control localities. No significant differences in breastfeeding outcomes were found. Any breast feeding at 6-8 weeks declined from 27% to 26% in intervention localities and increased from 29% to 30% in control localities (P=0.08, adjusted for pre-trial rate). Any breast feeding at 6-8 weeks increased from 38% to 39% in localities not participating in the trial. Women who attended breastfeeding groups were older (P<0.001) than women initiating breast feeding who did not attend and had higher income (P=0.02) than women in the control localities who attended postnatal groups. The locality cost was pound13 400 (euro14 410; $20 144) a year.

CONCLUSION: A policy for providing breastfeeding groups in relatively deprived areas of Scotland did not improve breastfeeding rates at 6-8 weeks. The costs of running groups would be similar to the costs of visiting women at home.

TRIAL REGISTRATION: Current Controlled Trials ISRCTN44857041.

3. **Pubmed ID: 11705561, Collected P-value(s): 0.03,0.05**

BACKGROUND: Why asthma is rare in rural subsistence societies is not clear. We tested the hypotheses that the risk of asthma is reduced by intestinal parasites or hepatitis A infection, and increased by exposure to dust-mite allergen or organophosphorus insecticides in urban and rural areas of Jimma, Ethiopia.

METHODS: From 12876 individuals who took part in a study of asthma and atopy in urban and rural Jimma in 1996, we identified all who reported wheeze in the previous 12 months, and a random subsample of controls. In 1999, we assessed parasites in faecal samples, Der p 1 levels in bedding, hepatitis A antibodies, serum cholinesterase (a marker of organophosphorus exposure), total and specific serum IgE, and skin sensitisation to Dermatophagoides pteronyssinus in 205 cases and 399 controls aged over 16 years. The effects of parasitosis, Der p 1 level, hepatitis A seropositivity, and cholinesterase concentration on risk of wheeze, and the role of IgE and skin sensitisation in these associations, were analysed by multiple logistic regression.

FINDINGS: The risk of wheeze was independently reduced by hookworm infection by an odds ratio of 0.48 (95% CI 0.24-0.93, p=0.03), increased in relation to Der p 1 level (odds ratio per quartile 1.26 [1.00-1.59], p=0.05), and was unrelated to hepatitis A seropositivity or cholinesterase concentration. In the urban population, D pteronyssinus skin sensitisation was more strongly related to wheeze (9.45 [5.03-17.75]) than in the rural areas (1.95 [0.58-6.61], p for interaction=0.017), where D pteronyssinus sensitisation was common, but unrelated to wheeze in the presence of high-intensity parasite infection.

INTERPRETATION: High degrees of parasite infection might prevent asthma symptoms in atopic individuals.

4. **Pubmed ID: 11907289, Collected P-value(s): 0.001**

BACKGROUND: Endothelin-1 is a potent vasoconstrictor and smooth-muscle mitogen. In a preliminary study, the orally administered dual endothelin-receptor antagonist bosentan improved exercise capacity and cardiopulmonary hemodynamics in patients with pulmonary arterial hypertension. The present trial investigated the effect of bosentan on exercise capacity in a larger number of patients and compared two doses.

METHODS: In this double-blind, placebo-controlled study, we randomly assigned 213 patients with pulmonary arterial hypertension (primary or associated with connective-tissue disease) to receive placebo or to receive 62.5 mg of bosentan twice daily for 4 weeks followed by either of two doses of bosentan (125 or 250 mg twice daily) for a minimum of 12 weeks. The primary end point was the degree of change in exercise capacity. Secondary end points included the change in the Borg dyspnea index, the change in the World Health Organization (WHO) functional class, and the time to clinical worsening.

RESULTS: At week 16, patients treated with bosentan had an improved six-minute walking distance; the mean difference between the placebo group and the combined bosentan groups was 44 m (95 percent confidence interval, 21 to 67; P<0.001). Bosentan also improved the Borg dyspnea index and WHO functional class and increased the time to clinical worsening.

CONCLUSIONS: The endothelin-receptor antagonist bosentan is beneficial in patients with pulmonary arterial hypertension and is well tolerated at a dose of 125 mg twice daily. Endothelin-receptor antagonism with oral bosentan is an effective approach to therapy for pulmonary arterial hypertension.

5. **Pubmed ID: 19741227, Collected P-value(s): 0.02,0.02**
BACKGROUND: Progressive multifocal leukoencephalopathy (PML) occurs in a fraction of patients with multiple sclerosis who were treated with natalizumab. Most adults who are infected with the JC virus, the etiologic agent in PML, do not have symptoms. We sought to determine whether exposure to natalizumab causes subclinical reactivation and neurotropic transformation of JC virus.

   METHODS: We followed 19 consecutive patients with multiple sclerosis who were treated with natalizumab over an 18-month period, performing quantitative polymerase-chain-reaction assays in blood and urine for JC virus reactivation; BK virus, a JC virus-related polyomavirus, was used as a control. We determined JC virus-specific T-cell responses by means of an enzyme-linked immunospot assay and antibody responses by means of an enzyme-linked immunosorbent assay and analyzed JC virus regulatory-region sequences.

   RESULTS: After 12 months of natalizumab therapy, the prevalence of JC virus in the urine of the 19 patients increased from a baseline value of 19% to 63% (P=0.02). After 18 months of treatment, JC virus was detectable in 3 of 15 available plasma samples (20%) and in 9 of 15 available samples of peripheral-blood mononuclear cells (60%) (P=0.02). JC virus regulatory-region sequences in blood samples and in most of the urine samples were similar to those usually found in PML. Conversely, BK virus remained stable in urine and was undetectable in blood. The JC virus-specific cellular immune response dropped significantly between 6 and 12 months of treatment, and variations in the cellular immune response over time tended to be greater in patients in whom JC viremia developed. None of the patients had clinical or radiologic signs of PML.

   CONCLUSIONS: Subclinical reactivation of JC virus occurs frequently in natalizumab-treated patients with multiple sclerosis. Viral shedding is associated with a transient drop in the JC virus-specific cellular immune response.

6. **Pubmed ID: 12215131, Collected P-value(s): 0.001,0.02**
CONTEXT: Carpal tunnel syndrome (CTS) can be treated with nonsurgical or surgical options. However, there is no consensus on the most effective method of treatment.

   OBJECTIVE: To compare the short-term and long-term efficacy of splinting and surgery for relieving the symptoms of CTS.

DESIGN, SETTING, AND PATIENTS: A randomized controlled trial conducted from October 1998 to April 2000 at 13 neurological outpatient clinics in the Netherlands. A total of 176 patients with clinically and electrophysiologically confirmed idiopathic CTS were assigned to wrist splinting during the night for at least 6 weeks (89 patients) or open carpal tunnel release (87 patients); 147 patients (84%) completed the final follow-up assessment 18 months after randomization.

MAIN OUTCOME MEASURES: General improvement, number of nights waking up due to symptoms, and severity of symptoms.

RESULTS: In the intention-to-treat analyses, surgery was more effective than splinting on all outcome measures. The success rates (based on general improvement) after 3 months were 80% for the surgery group (62/78 patients) vs 54% for the splinting group (46/86 patients), which is a difference of 26% (95% confidence interval [CI], 12%-40%; P<.001). After 18 months, the success rates increased to 90% for the surgery group (61/68 patients) vs 75% for the splinting group (59/79 patients), which is a difference of 15% (95% CI, 3%-27%; P =.02). However, by that time 41% of patients (32/79) in the splint group had also received the surgery treatment.

CONCLUSION: Treatment with open carpal tunnel release surgery resulted in better outcomes than treatment with wrist splinting for patients with CTS.

7. **Pubmed ID: 12241661, Collected P-value(s): 0.0001**

Urotensin II has vasoconstrictive and negative inotropic effects, suggesting a possible role in circulatory regulation and pathophysiology of heart failure. We developed a sensitive specific RIA and measured plasma urotensin II in patients with heart failure and in controls. Plasma urotensin II was higher in heart failure patients (mean 3.9 pmol/L [SD 1.4]; than in controls (1.9 pmol/L [0.9]; $p < 0.0001$). The role of urotensin II in heart failure, however, has yet to be defined.

8. **Pubmed ID: 12181103, Collected P-value(s): 0.0001**
In 1993-1995, the authors evaluated risk factors for elevated blood and bone lead levels in 264 Boston, Massachusetts, area women previously selected for a case-control study of lead and hypertension. Bone lead was measured at the tibia and patella with K x-ray fluorescence. Blood lead was analyzed by graphite furnace atomic absorption. Participants were aged 46-74 years and had mean lead levels of 3 (standard deviation, 2) micro g/dl (blood), 13 (standard deviation, 9) micro g/g (tibia), and 17 (standard deviation, 11) micro g/g (patella). In multivariate linear regression models, use of postmenopausal estrogen (inverse) and alcohol intake (positive) were significantly associated with blood lead levels. Both bone lead measures were significantly and positively associated with blood lead but only among postmenopausal women not using estrogen; for example, an increase from the first to the fifth quintile of tibia lead level (19 micro g/g) was associated with a 1.7- micro g/dl increase in blood lead (p = 0.0001) in this group. Older age and lower parity were associated with higher tibia lead; only age was associated with patella lead. The observed interaction of bone lead with estrogen status in determining blood lead supports the hypothesis that increased bone resorption, as occurs postmenopausally because of decreased estrogen production, results in heightened release of bone lead stores into blood.

9. **Pubmed ID: 11943693, Collected P-value(s): 0.0001,0.02**
Isoflavones are naturally occurring selective estrogen receptor modulators, with potential bone protective effects. To study the relation between soy isoflavone intake and bone mineral density (BMD), the authors analyzed baseline data from the Study of Women's Health Across the Nation,

a US community-based cohort study of women aged 42-52 years. Their 1996-1997 analysis included African-American (n = 497), Caucasian (n = 1,003), Chinese (n = 200), and Japanese (n = 227) participants. Genistein and daidzein intakes were highly correlated (r = 0.98); therefore, analyses were conducted by using genistein. Median intakes of genistein (measured in micrograms/day) by African Americans and Caucasians were too low to pursue relational analyses further. For Chinese and Japanese women, median genistein intakes were 3,511 and 7,151 microg/day, respectively. Ethnic-specific, linear models were used to predict BMD as a function of energy-adjusted tertile of intake, controlled for relevant covariates. For Chinese women, no association between genistein and BMD was found. Premenopausal, but not perimenopausal, Japanese women whose intakes were greater had higher spine and femoral neck BMD. Adjusted mean spinal BMD of those in the highest tertile of intake was 7.7% greater than that of women in the lowest tertile ($p = 0.02$); femoral neck BMD was 12% greater in the highest versus the lowest tertile ($p < 0.0001$).

10. **Pubmed ID: 14693873, Collected P-value(s): 0.0001,0.02,0.01,0.03,0.001**

CONTEXT: Complicated, left-sided native valve endocarditis causes significant morbidity and mortality in adults. The presumed benefits of valve surgery remain unproven due to lack of randomized controlled trials.

OBJECTIVE: To determine whether valve surgery is associated with reduced mortality in adults with complicated, left-sided native valve endocarditis.

DESIGN AND SETTING: Retrospective, observational cohort study conducted from January 1990 to January 2000 at 7 Connecticut hospitals. Propensity analyses were used to control for bias in treatment assignment and prognostic imbalances.

PATIENTS: Of the 513 adults with complicated, left-sided native valve endocarditis, 230 (45%) underwent valve surgery and 283 (55%) received medical therapy alone.

MAIN OUTCOME MEASURE: All-cause mortality at 6 months after baseline.

RESULTS: In the 6-month period after baseline, 131 patients (26%) died. In unadjusted analyses, valve surgery was associated with reduced mortality (16% vs 33%; hazard ratio [HR], 0.43; 95% confidence interval [CI], 0.29-0.63; $P < .001$). After adjustment for baseline variables associated with mortality (including hospital site, comorbidity, congestive heart failure, microbial etiology, immunocompromised state, abnormal mental status, and refractory infection), valve surgery remained associated with reduced mortality (adjusted HR, 0.35; 95% CI, 0.23-0.54; $P<.02$). In further analyses of 218 patients matched by propensity scores, valve surgery remained associated with reduced mortality (15% vs 28%; HR, 0.45; 95% CI, 0.23-0.86; $P = .01$). After additional adjustment for variables that contribute to heterogeneity and confounding within the propensity-matched group, surgical therapy remained significantly associated with a lower mortality (HR, 0.40; 95% CI, 0.18-0.91; $P = .03$). In this propensity-matched group, patients with moderate to severe congestive heart failure showed the greatest reduction in mortality with valve surgery (14% vs 51%; HR, 0.22; 95% CI, 0.09-0.53;$P = .001$).

CONCLUSIONS: Valve surgery for patients with complicated, left-sided native valve endocarditis was independently associated with reduced 6-month mortality after adjustment for both baseline variables associated with the propensity to undergo valve surgery and baseline variables associated with mortality. The reduced mortality was particularly evident among patients with moderate to severe congestive heart failure.

References

Allison, D. B., Gadbury, G. L., Heo, M., Fernandez, J. R., Prolla, T. A., and Weindruch, R. (2002). A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics and Data Analysis* **39**, 1-20.

Casella, G., and Berger, R. L. (2002). *Statistical Inference*, 2nd edition. Pacific Grove, CA: Duxbury.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1-38.

Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, 1st edition. New York, NY: Chapman and Hall.

Efron, B., Tibshirani, R. (2002). Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology* **23**, 70-86.

Genovese, C. R., Lazar, N. A., Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**, 870-8.

Kleinbaum, D. G. and Klein, M. (2005). *Survival Analysis: A self learning text*, 2nd edition. New York, NY: Springer.

Lehmann, E. L. (1991). *Testing Statistical Hypotheses*, 2nd edition. New York, NY: Springer.

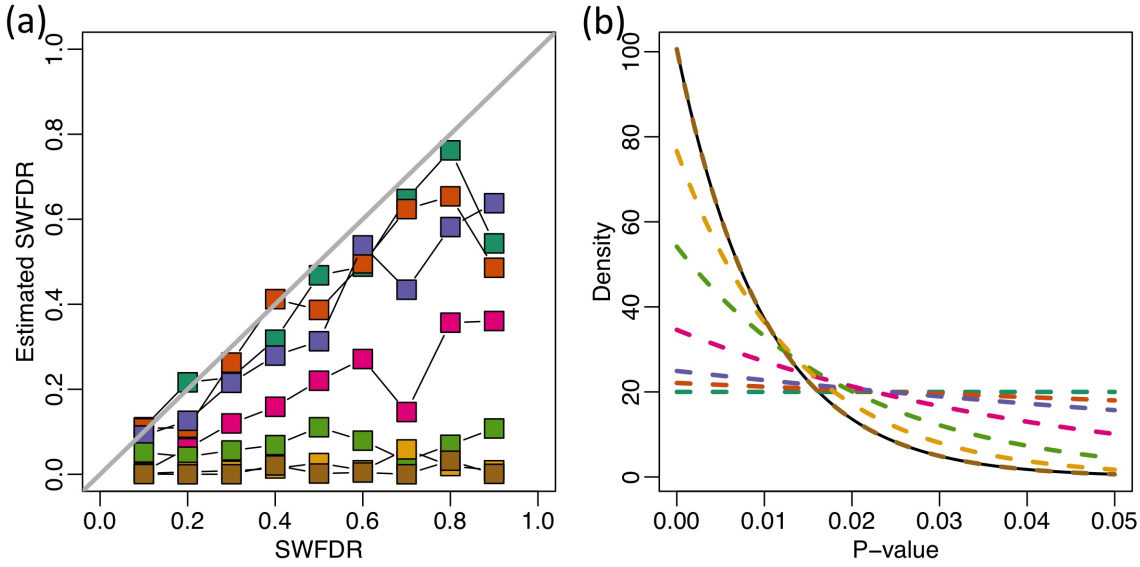
MacDonald, P. D. M. and Pitcher, T. J. (1979). Age-groups from size-frequency data: a versatile and efficient method of analysing distribution mixtures. *Journal of Fisheries Research Board of Canada* **36**, 987-1001.

Newton, M. A., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5**, 155-176.

STOREY, J. D., TIBSHIRANI, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440-5.

WENGRZIK, J. W. AND TIMM, J. (2011). Comparing Several Methods to Fit Finite Mixture Models to Grouped Data by the EM Algorithm. *Proceedings of the World Congress on Engineering* **1**, 1.

[*Received January 26, 2013; revised January 28, 2013*]



Supplementary Figure 1. **As the null distribution approaches the alternative distribution the science-wise false discovery rate estimates become anti-conservative.** (a) The science-wise false discovery rate (x-axis) versus estimated science-wise false discovery rate (y-axis) for varying null distributions. (b) A plot of the null distribution (dotted lines) where the color for the distribution matches the color for the corresponding false discovery rate estimates in panel (a). The line in black is the fixed alternative distribution for this figure.