

Oktober 2006

Volume 6 Nomor 1
ISSN. 0215.1995
Akreditasi Dirjen Dikti
No. 56/Dikti/Kep/2005

MAJALAH ILMU FAAL INDONESIA



The Indonesian Journal of Physiology

MIFI	Vol. 6	No. 1	Hal 1 - 73	Surabaya Oktober 2006	ISSN 0215 1995	Akreditasi 56/Dikti/Kep/2005
------	--------	-------	---------------	--------------------------	-------------------	---------------------------------

MAJALAH ILMU FAAL INDONESIA

Terbit setiap 4 bulan, distribusi seluruh Indonesia
Berdasarkan SK Ketua Umum Pengurus Pusat
Ikatan Ahli Ilmu Faal Indonesia
No. 001/SK/IAIFI/IX/2000 tanggal 14 September 2000



Penasihat : Ketua IAIFI Pusat
Penanggung Jawab : Dr. dr. Harjanto J. M., PFK (FK Unair)
Wakil Penanggung Jawab : dr. Choesnan Effendi, PFK (FK Unair)
Ketua Penyunting Pelaksana: dr. Adrianta Surjadhana, PFK (FK Unair)
Anggota : 1. dr. Paul Samuel Poly (FK UWK)
2. drg. Jenny Sunariani, MS (FKG Unair)
3. Dr. dr. Sunarko Setyawan, MS (FK Unair)
4. Dr. Drh. Anwar Ma'ruf, M.Kes (FKH Unair)
5. drg. Aqsa Sjuhada, M.Kes (FKG Unair)

Mitra Bestari : 1. Prof. dr. R. Soedarso Djojonegoro, PFK (FK Unair)
2. Prof. I. B. Adnyana Manuaba, HF, Fip. S (FK Udayana)
3. Prof. Dr. dr. Hari Lasmono, MS (F Psikologi Ubaya)
4. dr. R.M. Tauhid al Amien, M.Sc, Dipl.Grad.HPEd (FK Unair)
5. dr. Cholil Munif, PFK (FK Unair)
6. Dr. dr. Paulus Lieben, MS (FK Unair)
7. drh. Ngakan Made Rai Widjaja, MS (FKH Unair)
8. dr. Ibnu Mas'ud, MS (FK Unibraw)
9. dr. Joy Rattu, MS., Ph.D (FK Unsrat)

Tim Administrasi : 1. Drh. Kuncoro Puguh Santoso, M.Kes (FKH Unair)
2. Drh. Ratna Damayanti, M.Kes (FKH Unair)
3. dr. Bambang Purwanto (FK Unair)

Alamat : Bagian Ilmu Faal
Fakultas Kedokteran Universitas Airlangga
Jl. Prof. Dr. Moestopo 47
Surabaya 60132
Telp. (031) 5023621
Fax. (031) 5022472
E-mail : mifi@telkom.net

No. Rekening a.n. Ratna Damayanti, drh.,M.Kes
Bank BCA KCP Dharmahusada Surabaya
No. 3880294830

Harga Berlangganan Rp. 15.000; per nomor

3

Oktober 2006

Volume 6 Nomor 1
ISSN. 0215.1995
Akreditasi Dirjen Dikti
No. 56/Dikti/Kep/2005

MAJALAH ILMU FAAL INDONESIA



The Indonesian Journal of Physiology

MIFI	Vol. 6	No. 1	Hal 1 - 73	Surabaya Oktober 2006	ISSN 0215.1995	Akreditasi 56/Dikti/Kep/2005
------	--------	-------	---------------	--------------------------	-------------------	---------------------------------

DAFTAR ISI	ii
Roles Of Autonomic Nerves In Heart Rate Regulation Muhammad Ridwan	1
Dehydroepiandrosterone (Dhea) And Intracrinology Paulus Liben	14
Beda Efek Pemberian Epinefrin Akut Dan Kronis Terhadap Mcv Pada Tikus Putih Jantan Agustina RM, Tjitra Wardani, Sunarko Setiawan	23
Pengaruh Latihan Aerobik Gelombang Progresif Terhadap Kadar Lemak Tubuh Dan Kebugaran Aerobik Pada Penderita Overweight Djoko Pekik Irianto	36
Senam Kombo Intensitas Sedang Dibanding Ringan Terhadap Indeks Kebugaran Pada Wanita Lilik H, Choesnan E, Wahyuningtyas P	46
Aspek Fisiologis Pertahanan Tubuh Udang Windu (<i>Penaeus Monodon</i>) Terinfeksi <i>White Spot Baculo Virus (Wsbv)</i> Yang Diberikan Protein Immunogenik Vp28 Eduardus Bimo A.H	55
Uji Signifikansi Statistik: Konstroversi Dan Usulan Solusi Sia Tjundjing	64

TANGGAL 26 NOV 2014
MENGESAHKAN
FOTO COPY SESUAI DENGAN ASLINYA
Fakultas Psikologi
Universitas Surabaya
Dekan

a.n. 

Prof. Dr. Yusti Probowati R.
NPK. 190020

UJI SIGNIFIKANSI STATISTIK: KONSTROVERSI DAN USULAN SOLUSI

Sia Tjundjing^{*)}

ABSTRACT

One of the most reliable but also the most questionable statistical technique is the statistical significance test. Its existence is often regarded as one of the basic parameters of research quality, especially in hypothetical testing. On the other hand, its existence also creates an upsurge in critiques and questions for its dependence to the sample size. Controversies about its usage and alternative solutions are discussed.

Keywords :significance, controversies and solution

PENDAHULUAN

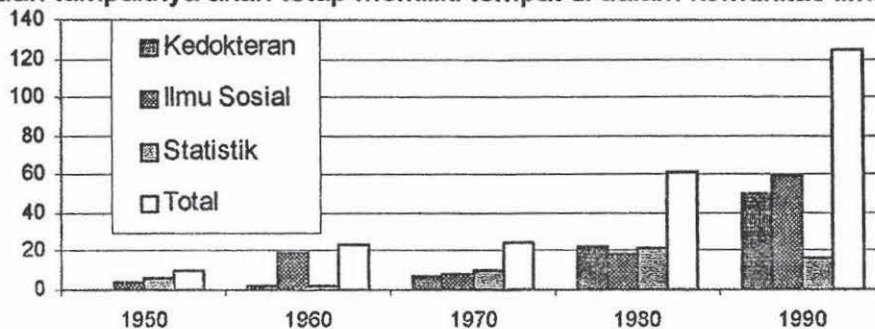
"Small wonder that students have trouble [with statistical hypothesis testing]. They may be trying to think" (Deming, 1975, p.150). Kutipan di atas hanyalah sekelumit dari komentar-komentar yang dilontarkan untuk mempertanyakan keabsahan pengujian signifikansi (*significance testing*), atau yang terkadang disebut juga *null hypothesis significance testing* (NHST) atau *statistical significance testing* (SST). Kritik tajam terhadap penggunaan uji signifikansi menjadi semakin semarak sejak artikel monumental dari Cohen (1994), "*The Earth Is Round (p < .05)*". Banyaknya perdebatan serta konvensi yang diadakan untuk membahas topik ini membuat *Board of Scientific Affairs* dari *American Psychological Association* (APA), membentuk tim kerja, *Task Force on Statistical Inference*, dengan melibatkan tokoh-tokoh terkemuka di bidang penelitian psikologi, statistik, dan ilmu sosial (Wilkinson & Task Force on Statistical Inference, 1999)

Hal ini merupakan fenomena menarik karena sekalipun telah diragukan dan dikritik oleh beberapa kalangan dari berbagai disiplin ilmu sejak 1980 (lihat Gambar 1), pelaporan hasil uji signifikansi tetap dilakukan secara konvensional. Hal ini juga dapat ditemui dalam jurnal ini. Kesalahan yang beberapa kali terjadi adalah pelaporan nilai p ($p < 0,05$ atau $p < 0,01$) tanpa disertai pelaporan ukuran efek (*effect size*), baik dalam bentuk koefisien korelasi atau indeks perbedaan, uji Anava/F atau Kai kuadrat/ χ^2 (Oki & Setyawan, 2005; Ma'ruf, 2005; Setiawan, 2004). Ada pula penulis yang sudah lebih akurat melaporkan ukuran efek namun kurang cermat dalam melaporkan nilai p dengan mengatakan bahwa nilai $p = 0,000$ atau $p = 1,000$ (Asnar & Kriswanto, 2005; Herawati, Effendi, & Puspitorini, 2004). Nilai p merupakan indeks

^{*)} Fakultas Psikologi, Universitas Surabaya

probabilitas, sehingga sangat jarang dan hampir tidak mungkin terjadi jika nilai p menunjukkan angka 0 atau 1 mutlak. Interpretasi yang lebih mungkin terjadi adalah nilai p tersebut terlampau besar (0,999987) atau terlampau kecil (0,0000123), kecuali apabila terjadi kesalahan perhitungan.

Tulisan kali ini berusaha memberikan penyegaran kembali konsep-konsep dasar maupun lanjut tentang seluk-beluk uji signifikansi. Hal ini dilakukan dengan mempertimbangkan bahwa sekalipun telah menuai banyak kritik sejak kemunculan pertama, uji signifikansi telah dan tampaknya akan tetap memiliki tempat di dalam komunitas ilmiah.



Gambar 1. Jumlah artikel yang meragukan kegunaan uji signifikansi dalam penelitian (dimodifikasi dari Buhi, 2005)

KESALAHPAHAMAN PENGERTIAN UJI SIGNIFIKANSI

Terdapat beberapa pengertian yang umumnya dikaitkan dengan definisi uji signifikansi. Berikut ini adalah beberapa pengertian yang biasa diberikan untuk menjelaskan nilai signifikansi yang diwakili dengan notasi " p " (Daniel, 1998).

1. Signifikansi statistik diartikan sebagai sesuatu yang penting, bermakna, atau signifikan.
2. Signifikansi statistik menunjukkan kebolehjadian (*likelihood*) bahwa hasil yang diperoleh akan tereplikasi di masa depan. Probabilitas memperoleh hasil yang sama apabila dilakukan penelitian ulang di masa mendatang yang diwakili oleh rumus $1-p$.
3. Nilai " p " menunjukkan kebolehjadian (*likelihood*) bahwa hasil yang diperoleh disebabkan oleh faktor kebetulan. Probabilitas hasil yang diperoleh hanya suatu kebetulan direfleksikan dengan nilai p .
4. Signifikansi statistik menunjukkan besarnya peluang bahwa sampel yang dipilih merupakan representasi dari populasi. Probabilitas bahwa sampel yang diambil merupakan representasi dari populasi diwakili dalam rumus $1-p$.
5. Signifikansi statistik dianggap sebagai solusi terbaik untuk mengevaluasi hasil uji atau analisis statistik.

6. Signifikansi statistik untuk suatu koefisien reliabilitas dan validitas dari sampel menunjukkan bahwa tes atau pengukuran tersebut juga akan reliabel dan valid apabila diberikan kepada kelompok sampel yang lain.

Sebelum dijelaskan secara lebih rinci pada bagian selanjutnya, ada baiknya melihat popularitas tiap-tiap jawaban di atas. Wright (2006) pernah mengajukan pertanyaan terkait uji signifikansi: "Probabilitas apakah yang diwakili oleh nilai p pada uji signifikansi statistik?" Persentase jawaban responden dapat dilihat pada Tabel 1.

Tabel 1. Persentase Jawaban Terkait Pengertian Nilai p

Alternatif Jawaban	%
Nilai p adalah probabilitas bahwa hasil yang diperoleh adalah karena kebetulan, probabilitas bahwa hipotesis nihil (H_0) adalah benar.	80%
Nilai p adalah probabilitas bahwa hasil yang diperoleh tidak sekedar kebetulan, probabilitas bahwa hipotesis nihil (H_0) adalah salah.	10%
Nilai p adalah probabilitas memperoleh nilai yang sama ekstrem (atau lebih) dengan nilai yang diperoleh, jika hipotesis nihil (H_0) adalah benar.	5%
Nilai p adalah probabilitas bahwa hasil eksperimen dapat direplikasi ketika dilakukan untuk kedua kalinya.	0
Semua salah.	5%

NHST DAN SIGNIFIKANSI PRAKTIS

Kesalahan ini mungkin merupakan kesalahpengertian yang paling sering dimunculkan terkait uji hipotesis, baik disadari atau tidak. Istilah signifikansi secara statistik seringkali disamakan dengan signifikansi secara praktis atau klinis. Hal ini terjadi karena penekanan dan penghargaan yang berlebihan pada hasil uji hipotesis. Penggunaan kata signifikan tanpa disertai kata "secara statistik" telah menjadi akar pemicu kesalahan ini. Dalam keadaan yang lebih parah, sekalipun telah disertai dengan kata "secara statistik", pemahaman yang ditangkap pembaca tetap salah, dan dianggap sebagai signifikan secara praktis.

Dalam menanggapi keterkaitan signifikansi statistik dan praktis, Thompson (1999, p. 39), selaku editor jurnal *Educational and Psychological Measurement*, memberikan pernyataan sebagai berikut, "... kejadian-kejadian yang jarang terjadi secara tidaklah menarik secara

intrinsik, namun tidak berarti tidak penting. Sebaliknya, hal-hal yang sering terjadi tidak juga selalu penting dan bermakna secara praktis." Oleh karena itu, disarankan adanya informasi tambahan terkait analisis kekuatan (ukuran efek) untuk melengkapi informasi hasil pengukuran atau analisis data. Hal ini menjadi penting dengan semakin banyaknya pendukung yang mengakui bahwa penambahan jumlah sampel akan mempermudah pemerolehan simpulan bahwa terdapat perbedaan antara kelompok data, walaupun perbedaan antara keduanya hanya sekitar 0,20.

Pelaporan ukuran efek (*effect size*) maupun taraf signifikansi secara bersamaan merupakan suatu keharusan dan tidak dapat dipisahkan untuk memperoleh makna yang lebih akurat untuk memahami data yang ada. Pemaparan ukuran efek dan taraf signifikansi akan memudahkan para pengambil keputusan dalam menentukan putusan yang paling optimal berdasarkan situasi yang dihadapi.

NHST DAN REPLIKABILITAS

Bertentangan dengan keyakinan dan penilaian beberapa pihak, Carver dan Thompson mengatakan bahwa uji signifikansi statistik bukanlah alat yang tepat untuk melihat replikabilitas hasil penelitian (sitat dalam Daniel, 1998). Kendati demikian, Thompson pernah membahas suatu prosedur untuk mengestimasi replikabilitas hasil penelitian, yaitu dengan validasi silang (*cross validation*), dan metode *jackknife* serta *bootstrap*. Semua metode tersebut pada intinya mendasarkan diri pada pemisahan sampel serta berbagai perhitungan statistik terhadap estimator-estimator berbeda dari berbagai variasi konfigurasi dari sampel yang sama dalam suatu studi. Sekalipun menuai kritik dengan memiliki bias karena hanya berasal dari satu sampel yang sama, Ferrelli telah menunjukkan bahwa hasil dari sebuah analisis regresi majemuk sederhana dapat divalidasi silang dengan membagi sampel asal secara *random*, dan memprediksi skor variabel untuk satu bagian sampel, dengan menggunakan pembobotan dari satu bagian sampel yang lain. Daniel, demikian pula Daniel dan Tucker (sitat dalam Daniel, 1998) juga pernah menggunakan logika serupa untuk melihat generalisabilitas hasil penelitian dengan menggunakan prosedur *jackknife*, yang secara serupa juga ditemui dengan pendekatan *bootstrap*.

NHST DAN FAKTOR KEBETULAN

Diskusi ini muncul karena banyaknya kesalahpahaman dengan menganggap bahwa uji signifikansi bertujuan untuk melihat seberapa tinggi peluang bahwa hasil yang diperoleh adalah suatu kebetulan. Untuk penyegaran kembali, NHST memang mengukur probabilitas, namun bukan probabilitas hasil yang terjadi karena kebetulan. Uji signifikansi berusaha mengetahui probabilitas hipotesis nihil diterima pada suatu populasi, yaitu populasi yang menjadi sumber dari sejumlah n sampel yang diambil. Sebagai dampaknya, uji signifikansi menyajikan informasi terkait probabilitas bahwa hasil yang diperoleh adalah suatu kebetulan, jika dan hanya jika hipotesis nihil diterima, dengan penggunaan *random sampling* dan pengambilan sampel yang memadai. Sebagai contoh: sebuah koefisien korelasi $r = 0,40$ diperoleh ketika menganalisis korelasi tes X dan Y, dengan jumlah sampel 100 siswa. Hasil terbukti signifikan pada tingkat 5%. Simpulan yang dapat diambil adalah terdapat 95% tingkat kebolehjadian bahwa korelasi antara kedua tes pada tingkat populasi tidak nihil, dengan mengasumsikan bahwa sampel tersebut merupakan perwakilan yang representatif dari populasi. Simpulan yang salah terjadi ketika mengatakan bahwa: (a) terdapat tingkat kebolehjadian sebesar 95% bahwa korelasi pada tingkat populasi adalah $r = 0,40$, atau (b) bahwa terdapat tingkat kebolehjadian sebesar 5% bahwa hasil tersebut hanyalah suatu kebetulan belaka. Komentar ini dimunculkan oleh Carver (1978).

NHST DAN PENGAMBILAN SAMPEL

Kesalahpahaman lain adalah dengan menafsirkan bahwa tujuan uji signifikansi adalah melihat tingkat keterwakilan populasi dalam sampel yang diambil. Nilai $1-p$ yang besar menunjukkan bahwa sampel telah mewakili populasi. Keterwakilan populasi dalam sampel tidak dapat diuji dengan uji signifikansi, melainkan hanya dengan memastikan bahwa proses pemilihan sampel telah dilakukan dengan hati-hati dan benar. Terkait dengan hubungan sampel dan populasi, sebenarnya NHST menjawab pertanyaan sebagai berikut, "Apabila sampel telah mewakili populasi, seberapa besarkah peluang memperoleh hasil penelitian?"

NHST DAN EVALUASI HASIL PENELITIAN

Kekeliruan ini menyatakan bahwa cara yang benar dan terbaik untuk mengevaluasi hasil penelitian adalah menggunakan uji signifikansi. Hal ini sering ditemui bersamaan dengan kesalahpahaman yang menganggap nilai p mencerminkan tingkat kepentingan hasil

penelitian. Sekalipun demikian, kesalahpahaman dalam menilai p sebagai indikator evaluasi dapat dianggap satu langkah lebih parah. Sebagai ilustrasi, kesalahan dengan menganggap nilai p sebagai indikator kepentingan hasil terlihat dari pembahasan yang hanya dilakukan berdasarkan nilai p semata, walaupun telah dilakukan pelaporan ukuran efek. Kesalahpahaman ini menjadi semakin parah ketika peneliti sama sekali (koefisien korelasi ataupun indeks perbedaan, nilai r , t , atau F) dan hanya melaporkan tingkat signifikansi.

Solusi yang ditawarkan Thompson adalah menyajikan semua ukuran efek secara lengkap sehingga pembaca dapat melakukan penilaian berdasarkan konteks masing-masing yang mereka hadapi. Solusi kedua adalah menyajikan analisis yang menyatakan jumlah sampel optimal dan minimal agar hasil yang diperoleh menjadi memadai, serta sebaliknya, dibutuhkan ukuran efek seberapa agar hasil dapat diterima untuk jumlah sampel yang sudah dipilih. Sebagai tambahan, Cohen juga mengingatkan apabila hasil uji ternyata tidak signifikan secara statistik, harus dilakukan perhitungan analisis kekuatan statistik (*statistical power analyses*) agar penyimpulan hasil dapat lebih optimal (sitat dalam Daniel, 1998).

NHST DAN KARAKTERISTIK SKOR TES

Validitas dan reliabilitas adalah ciri-ciri dari skor tes atau data yang diperoleh dari suatu tes. Sekalipun demikian, beberapa kali masih terjadi penyimpangan yang mengatribusikan validitas dan reliabilitas menjadi milik tes, sehingga disebut sebagai tes yang valid dan reliabel (padahal seharusnya skor tes yang valid dan reliabel). Penyimpangan ini sering dikaitkan dengan uji signifikansi, yaitu penggunaan NHST untuk koefisien validitas dan reliabilitas. Hal ini sebenarnya tidaklah tepat, karena koefisien reliabilitas dan validitas menjadi tergantung pada sampel yang digunakan. Sampel yang besar akan membuat koefisien yang kecil menjadi tetap bermakna. Sampel yang kecil akan membuat koefisien yang sudah memadai menjadi terlihat tidak bermakna. Oleh karena itu, penggunaan uji signifikansi sepertinya tidak tepat untuk mengevaluasi aspek validitas dan reliabilitas tes (Daniel, 1998). Pada akhirnya, Carver secara spesifik menawarkan empat cara untuk mengurangi ketergantungan pada uji signifikansi, yaitu:

- 1) menekankan penggunaan kata "secara statistik" pada akhir kata uji signifikansi,
- 2) mendesak bahwa interpretasi dilakukan terhadap data terlebih dahulu baru melihat hasil uji signifikansi,

- 3) mendesak interpretasi terhadap ukuran efek, terlepas dari signifikan atau tidak,
- 4) meminta para editor jurnal ilmiah untuk menyampaikan pandangan mereka terkait uji signifikansi sebelum terpilih menjadi editor jurnal.

Levin (1998) menyatakan persetujuannya dengan penggunaan uji asumsi, namun menyarankan dilakukannya perbedaan antara uji signifikansi secara statistik dengan estimasi statistik untuk mengetahui tingkat signifikansi praktis. Levin menegaskan bahwa uji signifikansi hanya menunjukkan bahwa terdapat perbedaan, namun apakah perbedaan tersebut diterima atau tidak itu masalah lain (terkait kesepakatan dalam bidang ilmu masing-masing terkait ukuran efek yang dianggap memadai).

USULAN PENGGANTI DAN PELENGKAP UJI SIGNIFIKANSI

Berikut ini adalah beberapa rekomendasi bagi para peneliti dalam menyikapi dan menggunakan uji signifikansi.

1. Selalu menggunakan kata "secara statistik" ketika mengacu pada hasil uji signifikansi. Tanpa menggunakan keterangan tersebut, hasil penelitian berpotensi disalahartikan.
2. Peneliti sebaiknya melaporkan ukuran efek (koefisien korelasi, nilai F , atau nilai t) sebagai pelengkap pelaporan uji signifikansi. Selain ukuran efek, statistik deskriptif terkait (parameter-parameter tendensi sentral seperti *mean*, *median*, dan *modus*) juga dapat dilaporkan sebagai pelengkap. Hal ini lebih tepat dipandang sebagai keharusan daripada saran. Pelaporan ini menjadi penting untuk memudahkan pembaca menilai dan menentukan sikap terhadap hasil penelitian. Hal ini tampaknya mulai menjadi budaya di antara para pengelola jurnal ilmiah, khususnya para pengelola jurnal di bawah naungan APA (*American Psychological Association*).
3. Rekomendasi berikutnya adalah pelaporan interval kepercayaan dan estimasi titik (*point estimate*). Interval kepercayaan akan menunjukkan interval kesalahan sekitar estimasi yang disusun dari sampel. Menurut Cohen (1990), minimnya pelaporan interval kepercayaan dikarenakan rasa malu peneliti ketika menyadari bahwa interval kepercayaan yang dihasilkan seringkali dinilai terlalu lebar. Interval kepercayaan yang lebar menunjukkan variasi data yang terlalu besar, secara sederhana, semakin kecil jarak antara interval kepercayaan, semakin terarah dan semakin baik pula hasil penelitian. Estimasi titik (*point estimation*) berguna untuk pengujian hipotesis terhadap suatu estimasi yang dapat dipertanggungjawabkan secara

- logis, dan tidak menggunakan angka nol sebagai tolok ukur, *nonzero estimate*.
4. Saran keempat mungkin terkesan berlebihan, namun akan sangat bermanfaat untuk dilakukan, yaitu melakukan meta-analisis. Meta-analisis akan menggabungkan hasil dari beberapa penelitian primer, dengan menggunakan beberapa ukuran efek dari setiap penelitian primer sebagai unit analisis. Sekalipun demikian, meta-analisis sangat rentan terhadap fenomena "*file drawer*" yang pernah disampaikan oleh Rosenthal (sitat dalam Daniel, 1998), yaitu tidak dilaporkannya penelitian dengan hasil yang tidak signifikan secara statistik sehingga hasil seringkali dianggap mengalami overestimasi. Meta-analisis memiliki keunggulan dalam hal menggabungkan hasil dari beragam penelitian dengan jumlah dan karakteristik subjek serta ukuran efek yang bervariasi.
 5. Peneliti sebaiknya melaporkan analisis terkait jumlah sampel, ukuran efek seberapa yang dibutuhkan agar hasil dianggap bermakna secara statistik, dan sebaliknya, dibutuhkan sampel sejumlah minimal berapa agar ukuran efek yang dihasilkan dapat dinilai signifikan secara statistik.
 6. Peneliti disarankan melaporkan replikabilitas eksternal dengan melakukan pengambilan ulang dari sampel yang berbeda atau melakukan analisis replikabilitas internal dengan menggunakan prosedur validasi silang, *jackknife* atau *bootstrap*.
 7. Melakukan analisis *power* sebagai pelengkap dari analisis signifikansi tradisional untuk memperkecil peluang kesalahan tipe dua (menyatakan hipotesis nihil diterima padahal sebenarnya hipotesis nihil ditolak)
 8. Teori keputusan. Terdapat kondisi-kondisi tertentu ketika hasil uji statistik tidak memadai untuk menentukan putusan terbaik yang harus diambil. Sebagai contoh, sekalipun diketahui bahwa suatu metode terapi lebih efektif dan efisien untuk meningkatkan kepercayaan diri, faktor biaya yang terlalu mahal dapat menyebabkan penolakan untuk menggunakan terapi tersebut. Dapat dilihat bahwa terdapat hal-hal lain yang mempengaruhi pengambilan keputusan, bukan hanya hasil uji statistik semata. Pada situasi inilah teori keputusan statistik sangat bermanfaat untuk menentukan kemungkinan keuntungan dan kerugian secara rasional di tengah ketidakpastian.
 9. Pemilihan model. Seiring dengan semakin maraknya penggunaan analisis statistik berbasis persamaan pemodelan struktural, semakin sering ditemui evaluasi terhadap model-model statistik tentatif. Thompson (1998) mengingatkan bahwa pemilihan dan penggunaan model "terbaik" secara statistik belum tentu paling baik pula secara

praktis. Dengan mempertimbangkan peluang adanya beberapa model yang tepat atau teruji, daripada memilih satu dan menggugurkan yang lain, apalagi hanya karena itu adalah model pertama yang "fit", disarankan untuk menyusun dan menguji semua model yang ditemui untuk selanjutnya dilakukan pemilihan, pembobotan, dan penerapan sesuai situasi dan kondisi yang dihadapi.

10. Pendekatan Bayesian. Dalam pendekatan Bayesian, diperkenalkan istilah interval kredibilitas (*credibility interval*). Berbeda dengan penggunaan interval kepercayaan, interval kredibilitas berusaha mengestimasi probabilitas nilai sebenarnya (*true value*) berada dalam interval kredibilitas (sesuai nilai alfa yang dihasilkan). Hal ini dinilai lebih bermanfaat daripada interval kepercayaan yang hanya dapat menyatakan bahwa apabila diulang secara berkali-kali sampai tak terhingga, terdapat kemungkinan sebesar 95% bahwa hasil penelitian terletak dalam interval kepercayaan yang diusulkan. Interval kepercayaan tidak dapat dan tidak berusaha mengestimasi probabilitas kesesuaian hasil penelitian dengan nilai sebenarnya pada populasi.

KESIMPULAN

Akhir kata, kemajuan di dunia ilmiah sama seperti halnya di bidang-bidang kehidupan lainnya, berada di tangan mereka yang terlibat di dalamnya. Oleh karena itu, untuk menanggapi berbagai saran dan kritik terkait uji signifikansi sikap yang tepat adalah bersikap terbuka dan tetap kritis untuk masukan-masukan yang diberikan. Secara umum, untuk meningkatkan kualitas penelitian, selain uji signifikansi secara statistik, diperlukan upaya-upaya untuk meningkatkan dan melaporkan ukuran efek dalam setiap penelitian serta mengupayakan replikabilitas hasil penelitian pada konteks penelitian yang berbeda.

DAFTAR PUSTAKA

- Asnar, E.S.T.P. & Kriswanto, E.S. (2005). Efek pemberian biji jarak merah (*Ricinus communis*) pada maturasi ovum mencit. *Majalah Ilmu Faal Indonesia*, (5), 1, 42-50.
- Bakan D (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.
- Buhi, E. R. (2005). The insignificance of "significance" tests: Three recommendations for health education researchers. *American Journal of Health Education*, 36(2), 109-112
- Carver, R.P. (1978). The case against statistical testing. *Harvard Educational Review*, 48, 378-399.

- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Daniel, L.G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the School*, 5(2), 23-32.
- Deming, W.E. (1975). On probability as a basis for action. *American Statistician* 29, 146-152.
- Herawati, L. Effendi, C., & Puspitorini, W. (2004). Komparasi latihan fisik kontinyu dengan interval terhadap glukosa darah segera dan 2 jam post-prandial. *Majalah Ilmu Faal Indonesia*, (4), 1, 23-32.
- Levin, J. R. (1998). What if there were no more bickering about significance tests? *Research in the Schools*, 5(2), 43-53.
- Ma'ruf, A. (2005). Peran fisiologis sekresi leptin sebagai dasar pencegahan obesitas. *Majalah Ilmu Faal Indonesia*, (5), 1, 18-23.
- Oki, A.S. & Setyawan, S. (2005). Pengaruh asam askorbat terhadap kadar imunoglobulin A sekretori pada saliva. *Majalah Ilmu Faal Indonesia*, (5), 1, 11-17.
- Setiawan, B. (2004). Efek pemberian laktoferin terhadap lebar vili usus halus pasca enterektomi ekstensif. *Majalah Ilmu Faal Indonesia*, (4), 1, 11-15.
- Thompson, B. (1999). Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. *The Journal of Psychology*, 133(2), 133-140.
- Thompson B. (1998). Five methodology errors in educational research: the pantheon of statistical significance and other faux pas. Paper presented at: Annual Meeting of the American Educational Research Association; April 15, 1998; San Diego, California.
- Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604
- Wright, D. (2006). Null hypothesis significance testing. Lecture 11. Retrieved May 7, 2006, from [http://www.sussex.ac.uk/Users/danw/masters/statistical analysis/lec11 NHST.ppt](http://www.sussex.ac.uk/Users/danw/masters/statistical%20analysis/lec11%20NHST.ppt)