Metodologi Penelitian dalam Psikologi Politik

Rizqy Amelia Zein1,2

1 Departemen Psikologi Kepribadian dan Sosial, Fakultas Psikologi Universitas Airlangga

2 Institute for Globally Distributed Open Research and Education (IGDORE)

Author note

Correspondence concerning this article should be addressed to Rizqy Amelia Zein, Kampus B Universitas Airlangga, Jalan Airlangga 4-6 Surabaya, Jawa Timur 60286. E-mail: [amelia.zein@psikologi.unair.ac.id](mailto:amelia.zein@psikologi.unair.ac.id)

Metodologi Penelitian dalam Psikologi Politik

Meskipun hitung cepat (*quick count*) mayoritas lembaga survei menunjukkan temuan yang bertolak belakang, Prabowo Subianto, calon Presiden pada Pemilihan Umum (Pemilu) 2019, buru-buru mengumumkan kemenangannya di depan awak media beberapa jam setelah Tempat Pemungutan Suara (TPS) ditutup. Berbekal temuan hitung cepat dan *exit poll* dari beberapa lembaga survei sekaligus perhitungan riil (*real count*) yang dilakukan oleh relawannya, Prabowo mengumumkan setidaknya ia meraup suara dengan margin yang sangat besar. Meskipun temuan lembaga survei yang dirujuk Prabowo bertolak belakang dengan kebanyakan lembaga survei termasuk perhitungan resmi Komisi Pemilihan Umum (KPU), Prabowo dan pendukungnya amat mempercayai temuan lembaga survei yang memenangkan Prabowo dan dengan lantang menuduh lembaga survei yang berbeda dengan klaimnya merupakan lembaga bayaran, tidak saintifik, dan partisan.

Dari kasus diatas tentu menarik untuk membahas pertanyaan-pertanyaan seperti; bagaimana sebenarnya cara terbaik untuk membedakan informasi yang saintifik, sains semu (*pseudoscience*), mis/disinformasi, dan khayalan? Bagaimana sesungguhnya cara kerja ilmuwan dalam menghasilkan temuan yang saintifik? Dan bagaimana seharusnya temuan penelitian dilaporkan dan disebarluaskan?

## 1. Logika saintifik (*scientific reasoning*)

Apabila anda mengingat kembali berbagai konsep yang telah anda pelajari di mata kuliah Pengantar Psikologi, anda mendapati hal-hal yang menarik, seperti; anak-anak dapat menirukan perilaku kekerasan (yaitu dengan memukul boneka Bobo) dengan mengamati orang dewasa yang melakukan perilaku tersebut, atau pembalap sepeda meraih waktu tempuh yang lebih cepat ketika membalap bersama orang lain daripada mengendarai sepedanya sendirian. Atau anda mungkin masih mengingat studi Latane dan Darley yang sangat populer, yang menunjukkan bahwa semakin banyak orang mendapati keadaan darurat, dimana seorang korban membutuhkan bantuan, justru menurunkan peluang korban tersebut mendapatkan bantuan. Mungkin anda juga bertanya-tanya mengapa anda tidak diajarkan cara membaca karakter seseorang dari garis tangan atau sidik jarinya, meskipun teknik ini sangat populer di kalangan awam.

Teori-teori di Psikologi pada dasarnya dibangun dengan pengamatan yang cermat dan sistematis, bukan yang ceroboh dan mengandalkan intuisi. Ilmuwan Psikologi tidak boleh menebak-nebak atau mencocok-cocokkan antara satu kejadian dengan kejadian yang lain. Untuk mengetahui apa yang bisa dan tidak bisa dipercaya, ilmuwan Psikologi sangat mengandalkan pengamatan yang cermat. Meskipun ada beberapa pendekatan yang menawarkan cara alternatif dalam merumuskan pengetahuan, pendekatan *empirisme* adalah yang paling dominan di sejarah perkembangan ilmu Psikologi.

Persoalan membedakan antara yang saintifik dengan yang tidak merupakan problem klasik yang diperdebatkan para filosof sejak dulu. Karl Popper (1902-1994) merupakan salah satu filosof yang cukup cerdik mengartikulasikan permasalahan ini dengan menyebutnya **problem demarkasi**. Popper sebenarnya tidak terlalu tertarik untuk memberikan label atas “mana yang saintifik dan yang tidak”, namun ia lebih banyak menjelaskan bagaimana strategi terbaik untuk menghasilkan pengetahuan yang saintifik.

Awalnya, filosof positivis percaya bahwa satu-satunya cara yang ampuh dalam memisahkan antara yang saintifik dengan yang tidak adalah dengan melakukan *verifikasi*. Prosedur ini dilakukan peneliti dengan memeriksa kesesuaian antara asumsi, teori, atau hipotesis dengan bukti empirik. Apabila bukti empirik yang tersedia konsisten dengan pernyataan yang hendak diuji, maka pernyataan tersebut dinyatakan terbukti sehingga dianggap saintifik.

Sebelum dicari korespondensinya pada realitas yang sedang diamati, pernyataan yang akan diuji harus memiliki **kriteria verifikasi** yang artinya, suatu teori **hanya dapat diverifikasi** ketika peneliti dapat menjelaskan secara rinci bagaimana langkah-langkah pengujiannya dalam bentuk **definisi operasional**. Definisi operasional adalah deskripsi yang memuat prosedur rinci yang harus dilakukan pengamat dalam menentukan eksistensi (ada/tidaknya) dan kualitas fenomena atau konsep yang diamati (Dienes, 2008).

Filosof yang mendukung positivisme tidak menyadari bahwa ada problem yang amat mendasar dari prosedur verifikasi. Utamanya ketika memastikan apakah bukti empirik yang diperoleh dari satu pengamatan akan berlaku pada pengamatan yang lain. Misalnya, ketika Ngadiman mengetahui bahwa (1) cabai berwarna merah, apakah (2) semua cabai akan memiliki warna yang sama? Pernyataan (1) dapat diperiksa melalui pengamatan langsung dengan logika **deduksi**, sedangkan permasalahan pada pernyataan (2) dapat diselesaikan dengan logika **induksi**.

**Deduksi** merupakan proses penarikan kesimpulan yang mengandaikan apabila terdapat hubungan yang logis antara dua pernyataan, maka kesimpulan yang mengikutinya akan sahih pula. Misalnya,

1. Semua cabai berwarna merah
2. Ngadiman memetik cabai

**Maka: Cabai Ngadiman berwarna merah**

Sedangkan induksi merupakan proses penarikan kesimpulan dari pengamatan berulang atas suatu gejala, misalnya

1. Ngadiman memetik cabai berwarna merah
2. Yuli memetik cabai berwarna merah
3. Sunu memetik cabai berwarna merah
4. Bejo memetik cabai berwarna merah
5. Dono memetik cabai berwarna merah

**Maka: semua cabai berwarna merah**

Kesimpulan dari logika induksi di atas tentu tak bisa serta-merta dianggap benar, sedangkan kesimpulan yang ditarik dengan logika deduksi juga sangat dangkal. Bagaimana mungkin dengan hanya satu kali mengamati, Ngadiman dapat yakin bahwa teorinya (semua cabai berwarna merah) tepat? Di sisi lain, meskipun Ngadiman tidak lagi sendirian mengamati warna cabai, rasanya janggal apabila kita langsung meyakini bahwa semua cabai berwarna merah. Namun tentu saja, apabila kita terus-menerus mendapati bukti yang konsisten setelah berulang kali mengamati, kita akan lebih percaya diri dalam menarik kesimpulan. Oleh karena itu, dapat disimpulkan bahwa kemungkinan kesimpulan pengamatan berlaku secara umum (generalisasi) akan meningkat seiring dengan ditemukannya bukti yang mendukung, ketika mengamatinya berulang-ulang.

Manariknya, Popper menolak asumsi ini. Klaim bahwa suatu pernyataan dapat dikatakan benar ketika bukti yang konsisten terakumulasi dari pengamatan berulang adalah sesuatu yang tidak masuk akal. Ketika kita akan mengamati warna cabai, tidak ada jaminan bahwa kita akan mendapati warna yang sama dengan pengamatan kita sebelumnya. Kita juga tidak bisa memperkirakan seberapa mungkin temuan kita sebelumnya, akan kita temui lagi di pengamatan berikutnya. Yang terpenting dari kritik Popper pada filosof positivisme adalah ketika mengasumsikan suatu teori adalah benar, sesungguhnya secara paradoks kita sekaligus mengasumsikan bahwa teori tersebut salah. Oleh karena itu, **suatu teori hanya akan benar selama tidak ada bukti yang dapat menggugurkannya**. Pernyataan *semua cabai berwana merah* akan langsung gugur apabila seorang pengamat menemukan ada *satu cabai yang berwarna hijau*. Popper menyebutkan bahwa pengetahuan saintifik tidak mungkin didapatkan dengan sekadar mengamati gejala. Merumuskan teori sesungguhnya bukan tujuan sains, karena kepastian dan kebenaran mutlak dari suatu teori tak mungkin dapat dicapai. Popper juga menunjukkan bahaya *fallibilism*, yaitu amat mungkin hal-hal yang kita percayai justru sesungguhnya sangat keliru.

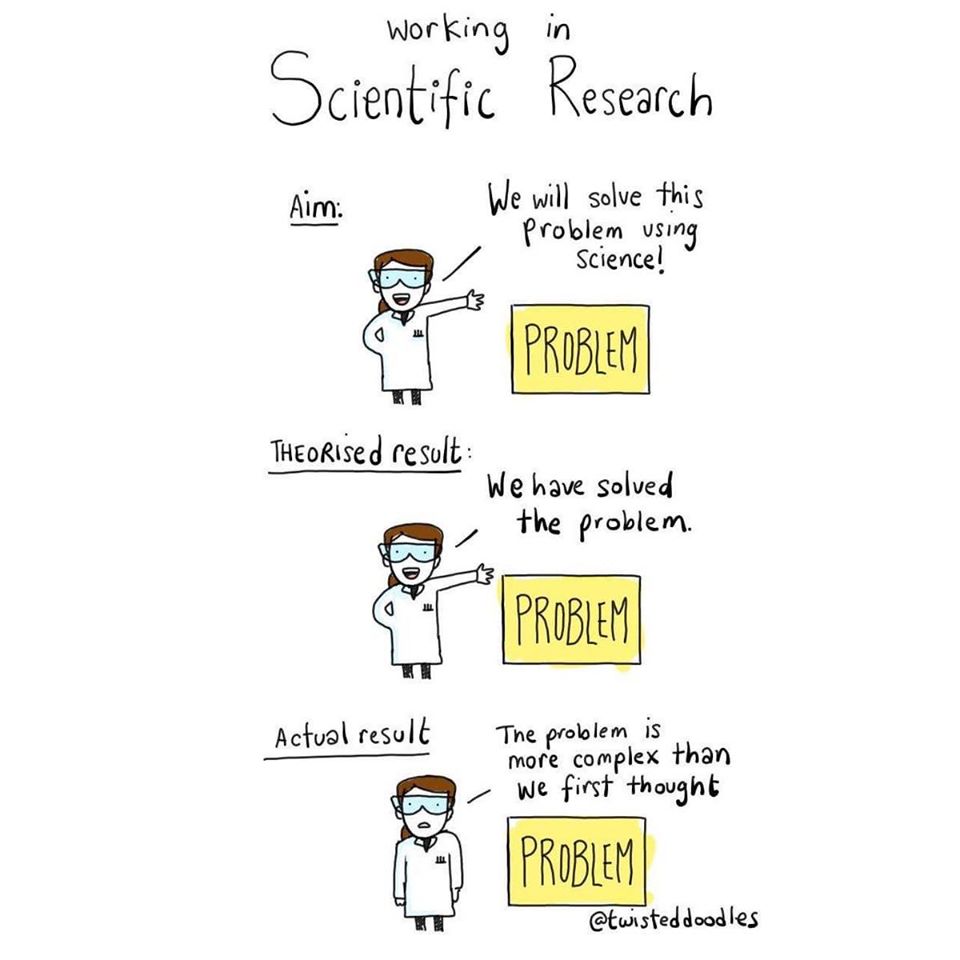
Sepanjang sejarah sains, kita sering menemukan episode dimana sekelompok orang berusaha mempertahankan tradisi mengenai ‘apa-apa yang sudah *kadung* mereka percayai’ sehingga penyimpangan atas tradisi ini berakibat pengucilan dan pembungkaman. Sebaliknya, sains justru berkembang dalam situasi yang penuh kontradiksi - ketika bukti baru, teknik baru, temuan baru, tokoh baru, menantang kepercayaan yang sudah mapan. Popper dengan tegas mengusulkan bahwa tujuan sains seharusnya adalah **memfalsifikasi**, yaitu mencari bukti yang bertentangan dengan keyakinan atau asumsi yang dimiliki oleh peneliti (Thornton, 2019). Apabila peneliti menemukan bukti mendukung teorinya, maka **bukan berarti** teori tersebut benar atau terbukti benar (*proved/established*), tetapi peneliti hanya dapat mengklaim bahwa teorinya **didukung oleh bukti** (*corroborated*). Teori tersebut berhasil dipertahankan dengan pengujian yang cermat melalui proses falsifikasi, bukan karena ia ‘mutlak benar’, namun karena ‘sementara ini’ teori tersebut yang paling mendekati kebenaran. Premis inilah yang kemudian menjadi sifat dasar sains, yaitu kesementaraan (provisional). Dengan kata lain, teori hanya dapat dianggap benar selama belum ada bukti yang menggugurkannya. Seiring dengan munculnya teknik baru, cara baru, dan bukti baru yang menambal kekurangan penelitian sebelumnya, maka pemahaman kita atas realitas akan semakin berkembang. Oleh karena itu, selain membutuhkan kebaruan dan inovasi, sains juga membutuhkan koreksi atas dirinya sendiri (*self-correction*) (Nosek et al., 2015).

## 2. Strategi pemeriksaan kredibilitas temuan penelitian

Sebagai mahasiswa Psikologi, tentu anda diwajibkan untuk mengerjakan proyek penelitian secara mandiri kemudian menulis laporan tugas akhir agar dapat dinyatakan lulus. Sebagian dari anda mungkin memiliki aspirasi untuk menjadi peneliti atau dosen di perguruan tinggi sebagai pilihan karir. Ada yang mungkin berkeingingan menjadi ilmuwan data (*data scientist*) yang banyak bersentuhan dengan *data mining* dan dataset skala besar untuk dianalisis. Dalam konteks ini, anda adalah *pelaku* penelitian - dimana anda banyak berperan dalam merumuskan pertanyaan penelitian, mendesain studi yang ditujukan untuk menjawab pertanyaan tersebut, melakukan pengambilan data, menganalisisnya, lalu melaporkan, dan mengkomunikasikan temuan pada khalayak yang lebih luas.

Sedangkan sebagian dari anda mungkin lebih tertarik menjadi praktisi - misalnya, bekerja sebagai manajer tim kampanye, atau sebagai konsultan politik. Sebagian mungkin tertarik bekerja di perusahaan media, atau mungkin bekerja sebagai tim ahli di parlemen atau di kantor pemerintah lainnya. Sebagian yang lain mungkin tertarik menjadi konselor bagi calon anggota dewan yang gagal terpilih. Dalam konteks ini, dalam mengerjakan pekerjaan anda sehari-hari, anda dituntut untuk mengambil keputusan yang **berbasis bukti** - yaitu, informasi yang didukung sehingga menjadi anda diharuskan menjadi *pengguna* hasil penelitian (Morling, 2018).

Pada praktiknya, anda melakukan kedua peran ini sehingga menjadi pengguna dan produsen riset sama pentingnya. Sebelum anda melakukan penelitian secara mandiri, anda harus belajar dari pengalaman peneliti yang telah bekerja sebelum anda. Anda perlu membaca literatur yang sesuai dan relevan dengan pertanyaan penelitian yang ingin anda ketahui. Melakoni kedua peran tersebut sama-sama membutuhkan rasa ingin tahu dan ketertarikan yang besar mengenai proses mental manusia. Menjadi pelaku sekaligus pengguna membutuhkan tidak hanya rasa ingin tahu, tetapi memiliki komitmen pada prinsip-prinsip dasar pemerolehan ilmu pengetahuan. Salah satu yang dapat dilakukan adalah berlatih untuk membaca literatur dan mengevaluasi informasi di dalamnya secara kritis.



*Figure* *1:*. Sumber: <https://twitter.com/twisteddoodles>

Dengan meminjam konsep falsifikasi Popperian, ada beberapa hal yang dapat dilakukan peneliti dalam mengevaluasi informasi dalam literatur secara kritis. *Yang pertama* adalah berulang kali mengajukan pertanyaan dan menghindari sikap ‘menerima segala hal begitu saja’ (*taken everything for granted*). Proses ini amat sulit pada awalnya dan membutuhkan latihan yang intensif. Namun yang perlu diingat, kemampuan ini amat sentral peranannya apabila kita ingin meninjau informasi secara kritis. *Yang kedua*, mengevaluasi semua bukti termasuk yang bertentangan dengan asumsi yang diyakini, untuk memahami ‘gambar besar’ dari fenomena yang sedang diselami. Ketika menarik kesimpulan, ilmuwan wajib mempertimbangkan seluruh bukti, bukan hanya yang sesuai dengan asumsi atau keinginannya saja. Melakukan yang terakhir dapat meningkatkan risiko terjadinya bias konfirmasi, dimana peneliti hanya percaya pada bukti-bukti yang sesuai dengan keinginannya saja. Apabila hal ini terjadi, maka sangat mungkin peneliti mendapatkan impresi yang keliru soal fenomena yang sedang ia pelajari.

*Yang ketiga*, ketika membaca hasil penelitian, cari tahu apakah studi tersebut dilandasi oleh teori yang jelas dan diikuti dengan prediksi yang jelas juga. Prediksi biasanya berbentuk hipotesis statistik yang spesifik, misalnya; ada korelasi yang positif/negatif antar variabel penelitian, ada perbedaan antara kelompok kontrol dengan *treatment*, dsb. Namun yang perlu dicatat, hipotesis statistik tidak secara langsung mencerminkan substansi dari teori, meskipun hipotesis tersebut diperoleh atau diturunkan dari teori (Dienes, 2008). *Yang keempat*, untuk melakukan pengujian apakah suatu teori sahih atau tidak, maka dibutuhkan hipotesis *auxilliary* (hipotesis minor) yang langsung dapat diujikan pada data. Misalnya, ketika anda ingin menguji kesahihan *dual-process motivational model of prejudice* (Duckitt, 2001) yang umum digunakan untuk menjelaskan munculnya prasangka pada kelompok sosial tertentu, maka anda perlu membuat hipotesis *auxilliary*, misalnya *“individu yang mempersepsikan dunia sebagai tempat yang berbahaya akan cenderung menyukai pemimpin yang otoriter”*. Dengan menggunakan hipotesis *auxilliary*, prediksi akan lebih mudah diuji dengan proses falsifikasi (*falsifiable*) yang dianjurkan oleh Popper.

Selain dapat difalsifikasi, teori atau prediksi harus dicari korespondensinya dengan realitas (data) melalui pengujian secara ketat. Pengujian ketat (*severe testing*) diperlukan untuk menyimpulkan apakah teori atau prediksi bertahan (*corroborated*) atau gagal dipertahankan (*falsified*). Sebuah pengujian dikatakan ketat (*severe*) apabila berdasarkan teori yang sedang diuji, kondisi yang diprediksikan sangat mungkin terjadi, namun kecil kemungkinannya terjadi berdasarkan semua teori lain yang tidak diuji (*the rest of background knowledge*) (Dienes, 2008).

Berikut adalah ilustrasi konkritnya.

Seorang peneliti ingin mengetahui apakah minum wedang jahe dapat meredakan masuk angin. Maka peneliti tersebut melakukan pengujian pada dua kelompok yang sama-sama menderita masuk angin - dimana satu kelompok diberikan segelas wedang jahe (kelompok *treatment*) sedangkan sisanya diberi minum air putih hangat (kelompok kontrol). Apabila kelompok yang diberi wedang jahe menunjukkan intensitas gejala masuk angin yang lebih rendah daripada kelompok kontrol, maka peneliti boleh menarik kesimpulan bahwa wedang jahe mungkin efektif mengobati masuk angin.

Namun prediksi ini juga dapat terjadi apabila kelompok *treatment* memiliki tingkat kekebalan tubuh yang lebih baik daripada kelompok kontrol. Oleh karena itu, penurunan gejala masuk angin dapat terjadi pada dua kondisi; (a) ketika minum wedang jahe; dan (b) ketika tingkat kekebalan tubuh tinggi. Kondisi ini adalah sesuatu yang **tidak diinginkan oleh peneliti**, karena pengujian ketat seharusnya mengasumsikan bahwa prediksi terjadi secara spesifik pada satu kondisi (teori yang sedang diuji) dan seharusnya tidak terjadi dalam kondisi (teori) yang lain. Oleh karena itu dalam penelitian tersebut, teori bahwa wedang jahe mengobati masuk angin tidak dapat difalsifikasi karena kita akan selalu mendapati hasil yang sama (penurunan gejala masuk angin) terlepas apakah prediksi tersebut (minum wedang jahe mengobati masuk angin) benar atau salah.

Dalam paradigma kuantitatif, utamanya yang menggunakan pendekatan Neyman-Pearson (Dienes, 2008; Perezgonzalez, 2015), pengujian ketat hanya mungkin dilakukan apabila peneliti memiliki peluang yang besar untuk mendeteksi bukti yang mendukung teori, apabila bukti tersebut benar-benar ada. Misalnya, ketika peneliti ingin tahu apakah minum wedang jahe benar-benar mengobati masuk angin, maka peneliti harus memiliki peluang yang besar untuk mendeteksi adanya perbedaan intensitas gejala masuk angin antara kelompok *treatment* dengan kelompok kontrol - kalau perbedaan tersebut benar-benar ada. Peluang untuk mendeteksi adanya efek (apabila efek tersebut benar-benar ada) dikenal sebagai ***statistical power*** (yang akan dijelaskan lebih lanjut di bagian selanjutnya). Apabila peneliti memiliki peluang yang kecil untuk mendeteksi efek (*statistical power* rendah), maka ketika peneliti tidak menemukan perbedaan intensitas gejala masuk angin antara kelompok *treatment* dengan kelompok kontrol, maka bukti tersebut tidak bisa menggugurkan (memfalsifikasi) teori bahwa wedang jahe menyembuhkan masuk angin. Oleh karena itu, dapat disimpulkan bahwa temuan penelitian dapat dianggap kredibel, apabila peneliti memiliki peluang (*statistical power*) yang besar untuk mendeteksi adanya efek, apabila efek tersebut benar-benar ada.

## 3. Isu mengenai pengukuran Psikologi

Pengukuran adalah bagian yang tidak mungkin dipisahkan dari proses penelitian. Tentu sulit bagi peneliti untuk menarik kesimpulan apakah dua kejadian berkaitan tanpa sebelumnya melakukan pengukuran pada dua kejadian tersebut. Oleh karena itu, penelitian yang baik tentu didahului oleh strategi pengukuran yang baik pula (Fried & Flake, 2018). Pengukuran dalam Psikologi pada prinsipnya sangat berbeda dengan mengukur gejala yang kasat mata, seperti berat badan, tinggi badan, atau warna rambut.

Sebelum melakukan pengukuran, kita perlu mengetahui perbedaan antara **konsep**, **konstruk**, dan **variabel**. **Konsep** merujuk pada deskripsi atas kejadian yang sifatnya umum dan merupakan kumulasi atau kumpulan dari beberapa konstruk. Sedangkan **konstruk** sendiri merujuk pada gejala yang sifatnya spesifik dan merupakan komponen kecil yang menjadi bagian dari konsep yang lebih besar. **Variabel** adalah *konstruk yang bervariasi*, sehingga anggaplah setelah dilakukan pengukuran pada satu kelompok sampel ditemukan tidak ada variasi konstruk dalam kelompok tersebut, maka **konstruk tidak dapat disebut sebagai variabel**.

Berikut ini adalah ilustrasi konkritnya.

Ilustrasi 1: Seorang Lurah ingin mengetahui gambaran kematangan fisik pada remaja di desanya dengan melakukan pengukuran tinggi dan berat badan. Setelah dilakukan pengukuran pada sampel sejumlah 100 remaja di desanya, maka ditemukan bahwa semua responden memiliki berat badan yang sama, yaitu 50kg, namun tinggi badannya bervariasi antara 140-156cm dengan rata-rata 152cm.

Ilustrasi 2: Seorang peneliti Psikologi Politik ingin menyelidiki persepsi pemilih mengenai calon Presiden dan Wakil Presiden dengan melakukan pengukuran pada tingkat keterpilihan pasangan calon Presiden dan Wakil Presiden yang sedang berlaga, dengan cara bertanya pada responden *“dari 0% (pasti tidak memilih) sampai 100% (pasti memilih), seberapa mungkin anda memilih pasangan A?”*. Setelah dilakukan pengukuran pada sekelompok sampel, pasangan A mendapatkan tingkat keterpilihan yang cenderung bervariasi antara 43-56% dengan rata-rata sebesar 45 persen.

Berdasarkan ilustrasi di atas maka **kematangan fisik** dan **persepsi pemilih** adalah **konsep**, sedangkan **tinggi, berat badan, dan tingkat keterpilihan** adalah **konstruk ukur**. Juga berdasarkan ilustrasi di atas, setelah dilakukan pengukuran maka diketahui **tinggi badan dan tingkat keterpilihan bervariasi** sehingga dapat disebut sebagai **variabel**. Namun **tidak ada variasi** pada **berat badan** di ilustrasi 1 sehingga meskipun berat badan merupakan konstruk ukur, **tidak dapat dikatakan sebagai variabel**.

Setelah membandingkan antara ilustrasi 1 dan 2, mungkin anda sudah mulai menangkap kesan mengenai kekhususan pengukuran Psikologi. Sebagian besar konstruk Psikologi **tidak dapat secara langsung diukur** melainkan diukur melalui manifestasi atau gejala yang dapat diamati secara langsung. Tentu kita dapat dengan mudah mengukur berat atau tinggi badan seseorang, namun mengukur konstruk Psikologi seperti intelegensi, persepsi, emosi, atau motivasi bukan pekerjaan mudah. Jangankan mengukur, mendefinisikan konstruk psikologi dengan lugas merupakan sesuatu yang sangat menantang. Sub-disiplin dalam Psikologi yang secara spesifik berfokus bagaimana mengkonstruksi dan mengevaluasi alat ukur Psikologi adalah Psikometri.

Dalam Psikometri, ada beberapa pendekatan yang umum digunakan untuk mengkonstruksi dan mengevaluasi alat ukur Psikologi. Diantaranya adalah pendekatan *latent variable modeling* (Skrondal & Rabe-Hesketh, 2007) dan *latent response model* (Maris, 1995). Namun pendekatan yang paling populer adalah Model Skor Klasik (*Classical True Score Model* atau CTM) (Mair, 2018). Berikut adalah persamaan dalam CTM:

Ide dasar dari CTT amat sederhana. Prinsipnya, *X* adalah skor yang langsung didapatkan dari suatu alat ukur (*observed score*) atau disebut juga sebagai skor kasar (*raw score*). Skor kasar mengandung dua elemen, yaitu skor murni (*true score*) atau *T* dan kesalahan pengukuran (*measurement error*) atau *e*. Yang menarik, *T* dan *e* tidak mungkin diketahui nilainya dengan pasti. Dengan begitu kita dapat menyimpulkan bahwa:

1. Tidak ada satupun alat ukur Psikologi yang mampu mengukur konstruk Psikologi dengan sempurna sehingga peneliti harus meminimalisasi dan menoleransi kesalahan pengukuran dalam kadar tertentu.
2. Ketika peneliti melakukan pengukuran berulang pada orang yang sama di waktu yang berbeda, maka tidak realistis apabila peneliti tersebut mengharapkan hasil pengukuran yang *identik* padahal konsistensi hasil pengukuran merupakan aspek penting untuk mengevaluasi apakah alat ukur tersebut bekerja dengan baik.
3. Yang paling krusial dari CTM adalah kesalahan pengukuran (*e*). Apabila kita tidak mungkin mengetahui nilai *e*, lalu bagaimana caranya mengetahui apakah alat ukur kita akurat atau tidak? Meskipun secara realitanya kita tidak mungkin mengetahui nilai *e*, namun karena nilai *e* dapat bervariasi, maka kita dapat mengestimasi variasi dari *e* dengan menghitung *standard error of measurement* (E). SoE dan reliabilitas (akan dijelaskan di bagian selanjutnya) merupakan indikator yang penting yang dapat digunakan peneliti untuk menimbang apakah pengukurannya akurat atau tidak.

Untuk mengetahui pengukuran yang dilakukan sudah tepat, kita dapat memastikannya melalui dua hal; (a) apakah alat ukur benar-benar mengukur objek yang ingin diukur (validitas), dan (b) dengan mengasumsikan tidak ada perubahan berarti pada konstruk yang diukur, apakah alat ukur akan memberikan hasil yang cenderung konsisten (reliabilitas). Dua fitur inilah yang kemudian menentukan apakah hasil pengukuran dari suatu alat ukur Psikologi dapat dipercaya.

## 3.1 Reliabilitas

Reliabilitas adalah indikasi kualitas alat ukur dalam konteks CTM yang juga merefleksikan keterkaitan antara *X*, *T*, dan *e*. Alat ukur yang reliabel akan memberikan hasil pengukfuran yang konsisten ketika ia diujikan pada orang yang sama pada waktu yang berbeda. Dengan kata lain, reliabilitas adalah indikator akurasi sebuah tes karena menunjukkan kemampuan alat ukur memberikan estimasi skor yang mendekati skor murni (Mair, 2018). Ada dua skenario yang mungkin terjadi - ketika alat ukur memiliki reliabilitas yang baik, maka varians kesalahan pengukuran (E) akan mengecil sehingga *X* memiliki kadar kesalahan pengukuran yang kecil dan makin mendekati nilai *T*. Kemungkinan yang kedua dan yang tak diinginkan peneliti adalah ketika alat ukur memiliki kadar reliabilitas yang buruk, maka E akan membesar sehingga *X* terkontaminasi kesalahan pengukuran yang besar sehingga nilai *X* makin menjauh dari *T*.

Untuk menakar kadar reliabilitas alat ukur, ada beberapa pendekatan yang dapat diadopsi oleh peneliti, diantaranya adalah: *test-retest*, tes paralel, dan konsistensi internal. Pendekatan yang pertama dapat dilakukan dengan mengujikan alat tes pada kelompok sampel yang sama, namun pada waktu yang berbeda, sedangkan yang kedua mengujikan alat ukur pada dua kelompok sampel yang berbeda, namun memiliki karakteristik yang identik (homogen). Baik *test-retest*, maupun tes paralel dianggap kurang praktis karena membutuhkan sedikitnya dua kali pengambilan data. Oleh karena itu, pendekatan yang terakhir (konsistensi internal) adalah yang paling sering digunakan.

Pendekatan konsistensi internal prinsipnya menguji reliabilitas akat ukur psikologi dengan melihat konsistensi antar butir pertanyaan/pernyataan (aitem). Pendekatan ini bercabang pada dua asumsi yang berbeda, yaitu peneliti dapat mengasumsikan bahwa kontribusi setiap aitem **sama besarnya** atau **bervariasi/berbeda-beda**  dalam menjelaskan konstruk yang diukur. Apabila kontribusi aitem **sama besarnya** dalam menjelaskan konstruk psikologis yang diukur, maka alat ukur tersebut memenuhi asumsi *Equivalence* sehingga teknik Cronbach’s dan *Split-Half* dapat digunakan untuk mengestimasi reliabilitas alat ukur (Cho, 2016). Namun yang perlu diingat, asumsi *Equivalence* **tidak realistis** karena pada kenyataannya, tidak mungkin peneliti mengharapkan bahwa seluruh aitem dalam alat ukurnya berkontribusi setara dalam menjelaskan konstruk. Menariknya, meskipun asumsi *Equivalence* tidak realistis, teknik Cronbach’s adalah pilihan yang sangat populer dan sayangnya, paling sering disalahgunakan oleh peneliti (McNeish, 2018). Ketika menggunakan Cronbach’s , peneliti seringkali abai pada asumsi yang menyertainya (Peters, 2018).

Alternatif lainnya adalah peneliti mengharapkan masing-masing aitem dalam alat ukurnya **berkontribusi secara berbeda** dalam menjelaskan konstruk. Asumsi ini lebih fleksibel dan lebih dekat dengan realitas yang dihadapi oleh peneliti daripada *Tau Equivalence*. Asumsi ini dikenal sebagai *model konjenerik* dan untuk mengestimasi reliabilitas alat ukur Psikologi dengan asumsi konjenerik, peneliti dapat menggunakan beberapa alternatif, yaitu *greatest lower bound* (glb) (Sijtsma, 2008) dan *McDonald’s Omega* () (Dunn, Baguley, & Brunsden, 2014). Apabila peneliti tertarik untuk mengestimasi reliabilitas pada alat ukur yang mengukur konstruk yang sifatnya unidimensional, maka Omega Total (t) adalah teknik yang terbaik. Namun untuk konstruk yang multi-dimensional atau memiliki struktur tertentu, maka pendekatan Omega Hirarkial (h) adalah yang terbaik (Revelle & Zinbarg, 2008). Program statistik dengan sumber terbuka seperti [jamovi](https://www.jamovi.org/), [JASP](http://jasp-stats.org/), atau R dengan [*package* psych](https://cran.r-project.org/web/packages/psych/index.html) dapat membantu peneliti untuk melakukan kalkulasi reliabilitas dengan teknik .

## 3.2 Validitas

Isu mengenai **validitas** alat ukur Psikologi berfokus untuk memastikan bahwa alat ukur benar-benar mengukur konstruk yang ingin diukur dan bukan konstruk lainnya. Apabila kita mengukur gejala tampak, mengevaluasi validitas alat ukur sangat mudah dilakukan. Misalnya, kita dapat dengan mudah menyimpulkan bahwa penggaris bukan alat ukur yang valid untuk mengukur berat badan. Namun bagaimana dengan konstruk Psikologi?

Pendekatan tradisional yang lazim digunakan untuk menyelidiki validitas alat ukur Psikologi membagi validitas berdasarkan tiga kategori, yaitu: validitas isi (*content validity*), validitas kriteria (*criterion validity*), dan validitas konstruk (*construct validity*) (Messick, 1995). Validitas isi dapat diperoleh dengan meminta bantuan ahli (*expert judgement*) untuk menilai apakah alat ukur sudah mencerminkan konstruk yang ingin diukur (Wynd, Schmidt, & Schaefer, 2003), sedangkan validitas kriteria dapat diperoleh dengan mengkorelasikan hasil pengukuran dengan kriteria lain yang relevan. Mengestimasi validitas konstruk dapat dilakukan dengan mengkorelasikan skor alat ukur yang akan diuji dengan konstruk lain yang dihipotesiskan berkorelasi linier dan positif (*convergent validity*) atau dengan konstruk lain yang dihipotesiskan sumbang (*divergent validity*) (Messick, 1995). Msisalnya apabila anda ingin mengestimasi validitas alat ukur yang mengukur depresi, maka anda akan mengkorelasikan skornya dengan kecemasan dan kebahagiaan. Asumsinya, apabila alat ukur anda valid, maka korelasinya dengan depresi akan kuat dan positif, sedangkan korelasi skor alat ukur anda akan cenderung negatif dengan kebahagiaan. Validitas kriteria juga dapat dicek dengan melakukan kedua teknik tersebut sekaligus (*convergent* dan *divergent*) dengan membuat matriks korelasi yang menggambarkan korelasi antara skor alat ukur yang diuji dengan alat ukur lain yang sama-sama mengukur *trait* atau konstruk yang sama (bisa juga dengan metode yang berbeda). Pendekatan ini populer dikenal dengan *Multitrait-Multimethod* (MTMM) (Dumenci, 2000).

Yang terakhir adalah validitas konstruk. Sebenarnya kedua tipe validitas yang diuraikan sebelumnya tidak cukup ampuh dan cermat untuk mengevaluasi validitas alat ukur Psikologi karena pada dasarnya apabila kita ingin tahu apakah alat tes Psikologi benar-benar mengukur konstruk yang ingin diukur, maka peneliti harus memastikan bahwa butir pertanyaan/pernyataan (aitem) mengukur konstruk laten (yang sama). Ini adalah definisi dari validitas konstruk (Cronbach & Meehl, 1955) dan mayoritas ahli Psikometri sepakat bahwa validitas alat ukur harus selalu merujuk pada validitas konstruk, bukan yang tipe validitas yang lain (Messick, 1995). Messick (1995) bahkan mengusulkan *unified validity* dimana pembagian tiga kategori validitas tidak perlu lagi dilakukan. Ketika seorang ingin tahu validitas suatu alat ukur Psikologi, maka langsung saja merujuk pada pendekatan validitas konstruk.

Ada beberapa alternatif pendekatan validitas alat ukur, yaitu: *incremental validity* yang berfokus pada pertanyaan: apakah alat ukur yang diujikan menambah pemahaman (atau kekuatan prediksi) peneliti dibandingkan dengan alat ukur lain yang sudah lebih dulu tersedia (Hunsley & Meyer, 2003). Pendekatan lainnya yang tersedia adalah *Processed-Focused Model* yang lebih banyak berkutat pada penjelasan mengenai keterlibatan partisipan tes dalam proses pengukuran dan apakah keterlibatan tersebut sudah sesuai dengan sifat dan situasi tes yang sudah diantisipasi sebelum tes berlangsung (Bornstein, 2011).

## 4. Ragam desain penelitian dalam studi-studi Psikologi Politik

Setelah peneliti memahami cara kerja sains dan mampu mendesain strategi pengukuran yang sesuai, maka peneliti dapat mulai memilih desain penelitian yang sesuai dengan pertanyaan penelitian yang ingin dijawab atau hipotesis yang akan diuji. Apabila menganalogikan kegiatan meneliti dengan memasak, maka memilih desain penelitian yang sesuai dengan pertanyaan penelitian seperti memilih alat yang digunakan untuk memasak. Untuk memilih alat masak yang tepat, maka kita harus benar-benar memahami apa yang ingin kita masak. Serta untuk memastikan masakan dapat dihidangkan sebagaimana mestinya, maka kita harus mengikuti tahapan memasak yang tepat. Melewati beberapa langkah atau tidak melakukan tahapan memasak dengan berurutan akan berdampak pada kualitas masakan.

Paralel dengan kegiatan meneliti - bahwa peneliti harus benar-benar memahami apa yang ia hendak cari dengan merumuskan pertanyaan penelitian dan hipotesis yang tepat. Agar temuan penelitian kredibel, maka peneliti perlu memilih desain yang paling sesuai dalam menjawab pertanyaan penelitian atau menguji hipotesis yang ia ajukan. Terakhir, apabila peneliti menginginkan temuan yang meyakinkan, maka ia perlu cermat dalam mengikuti tahapan atau prosedur yang sistematis dalam melakukan penelitian.

Di bagian ini, penulis akan menguraikan secara singkat beragam desain penelitian yang lazim digunakan dalam studi Psikologi Politik.

## 4.1 Survei

Bayangkan ada sebuah ceret yang berisi kopi di hadapan anda saat ini. Untuk mengetahui rasa kopi, apakah anda harus meminum seluruh kopi yang di dalam ceret tersebut? Kalau dengan mencicip sedikit saja sudah cukup, seberapa banyak kopi yang harus diminum agar dapat menyimpulkan rasanya dengan akurat? Logika dan persoalan yang sama juga dihadapi oleh peneliti Psikologi Politik yang menggunakan desain survei. Prinsipnya, untuk mendapatkan berbagai informasi mengenai sekelompok orang, kita tidak perlu melakukan pengujian pada seluruh anggota kelompok (populasi), namun cukup dengan melakukan pengujian pada beberapa orang yang mewakili kelompok tersebut (sampel).

Penelitian survei umumnya dapat berupa **survei *cross-sectional***, yaitu ketika peneliti melakukan pengambilan data pada satu kelompok sampel dalam satu waktu. Atau survei dapat dilakukan dengan desain **longitudinal** (*time series* atau *panel study*), yaitu ketika peneliti melakukan pengambilan data secara berulang pada kelompok sampel yang sama dalam periode waktu tertentu. Data yang diperoleh dari penelitian survei memiliki kegunaan praktis yang luas melampaui tujuan akademik, diantaranya sebagai dasar perumusan kebijakan dan pengambilan keputusan lainnya di berbagai konteks. Survei skala besar seperti Survei Sosial Ekonomi Nasional (SUSENAS) oleh Badan Pusat Statistik (BPS), Riset Kesehatan Dasar (RISKESDAS) oleh Badan Penelitian dan Pengembangan Kesehatan (Balitbangkes) Kementerian Kesehatan, atau Indonesian Family Life Survey (IFLS) oleh RAND, menyediakan data yang amat kaya dan memberikan banyak informasi penting yang dapat digunakan untuk mendukung perumusan kebijakan di berbagai sektor.

Penelitian survei merupakan desain yang amat efisien dan paling luas diaplikasikan pada penelitian sosial dan perilaku, termasuk dalam studi Psikologi Politik. Penelitian survei dapat dilakukan untuk kepentingan akademik dan juga lazim digunakan untuk memberikan informasi kepada pemilih mengenai kandidat yang akan mereka pilih di pemilihan umum. Di bagian ini, penulis akan lebih banyak menjelaskan mengenai survei dalam konteks studi akademik, meskipun di bagian akhir tersedia sedikit ulasan mengenai survei yang dilakukan oleh lembaga survei non-pemerintah merangkap penyedia jasa konsultan politik (*pollster*) utamanya yang menjadi bagian dari kegiatan pemasaran politik (*political marketing*).

**4.1.1 Populasi dan sampel**

Survei merupakan desain penelitian yang tak hanya dapat digunakan untuk membuat profil yang mendeskripsikan sekelompok orang, namun juga dapat digunakan untuk menguji teori atau prediksi. Namun seringkali peneliti memiliki sumberdaya yang terbatas sehingga tidak mungkin melibatkan seluruh anggota populasi (sensus) dalam pengambilan data. Yang juga sering terjadi adalah ukuran populasi yang tidak diketahui sehingga peneliti harus memilih sebagian kecil responden yang dapat mewakili populasi yang ingin diteliti. Tidak hanya itu, kualitas data yang diperoleh melalui proses *sampling* justru lebih baik daripada data sensus (Fowler, 2014). Proses pemilihan responden ini disebut dengan **teknik *sampling***.

Ada beberapa istilah yang perlu diketahui agar dapat melakukan proses pemilihan sampel dengan baik. **Populasi** adalah sekumpulan individu yang memiliki karakteristik atau kriteria tertentu yang telah ditetapkan oleh peneliti sebelum studi dilakukan. Populasi juga mencerminkan konteks yang ingin diteliti dan merupakan bagian penting dari studi yang dilakukan. Kriteria yang ditetapkan oleh peneliti untuk mendefinisikan populasi penelitiannya disebut juga **kriteria inklusi**. Populasi biasanya disimbolkan dengan *N* dan harus definisikan secara eksplisit, hati-hati, dan menyeluruh dengan memperhatikan pertanyaan penelitian yang ingin dijawab (Ruel, Wagner, & Gillespie, 2016).

Mendefinisikan populasi tidak semudah kelihatannya. Selain menentukan kriteria inklusi, peneliti juga perlu menerapkan **kriteria eksklusi** yang kemudian menghasilkan **populasi target**. **Kriteria eksklusi** adalah karakteristik yang mengecualikan individu tertentu dari proses pemilihan sampel. Pengecualian ini biasanya berkaitan dengan keterbatasan sumberdaya yang dimiliki oleh peneliti. Misalnya, apabila seorang peneliti ingin melakukan penelitian pada Warga Negara Indonesia (WNI), kadang-kadang peneliti harus mengecualikan WNI yang tinggal di luar negeri karena sumberdaya yang dimiliki peneliti tidak cukup untuk melibatkan WNI yang tinggal di luar negeri.

**Sampel** adalah subset atau bagian kecil dari populasi, yang biasanya disimbolkan dengan *n*. Elemen sampel dapat berbentuk **unit analisis** yang beragam, mulai dari individu, kelompok (perusahaan, ormas, universitas, dll.), objek, tempat (RT, RW, Kabupaten/Kota, Propinsi, Desa, dll.), atau kejadian yang menarik untuk diteliti. Apabila populasi penelitian terdiri dari sekumpulan individu, maka sampel juga terdiri dari individu-individu. Untuk menentukan anggota populasi yang menjadi sampel, peneliti perlu menjamin bahwa sampel tersebut benar-benar representatif mewakili populasi. Apabila sampel yang dipilih representatif mewakili populasi, maka temuan penelitian yang diperoleh dari proses *sampling* dapat **digeneralisasi** pada populasi yang diwakilinya. Untuk menjamin **generalisasi**, peneliti perlu memberikan kesempatan yang sama pada seluruh anggota populasi untuk terpilih menjadi sampel. Pendekatan ini disebut juga sebagai ***probability sampling***.

Bayangkan anda sedang membuat kopi dengan gula aren sebagai pemanis. Yang anda lakukan adalah menuang gula aren terlebih dahulu, baru menuang kopi yang sudah dicampur dengan sedikit air. Untuk mendapatkan rasa yang optimal, maka yang harus anda lakukan adalah **mengaduk** semua cairan yang ada di dalam gelas. Apabila anda tidak mengaduknya, maka kopi tidak terasa manis karena gula aren masih mengendap di bagian bawah gelas. Dengan mengaduk campuran di dalam gelas, anda mencampur dengan rata semua elemen yang ada di dalam gelas sehingga dengan mencicip sedikit, anda sudah dapat merasakan profil kopi secara keseluruhan. Namun apabila tidak diaduk, anda hanya akan merasakan rasa kopi dan meninggalkan rasa manis yang diberikan oleh gula aren. Cara kerja *probability sampling* persis seperti analogi pengadukan ini. Mengaduk cairan yang ada di dalam gelas berfungsi sebagai **pengacakan**. Dengan melakukan **pengacakan**, berarti anda memberikan **kesempatan yang sama** pada seluruh anggota populasi untuk terpilih menjadi sampel. Dengan begitu, anda dapat lebih percaya diri untuk menyimpulkan bahwa sampel anda representatif sehingga dapat dilakukan generalisasi.

Untuk melakukan *probability sampling*, maka peneliti harus memiliki ***sampling frame***, yaitu daftar yang memuat seluruh anggota populasi. Apabila peneliti tertarik melakukan studi dengan populasi WNI, maka untuk melakukan *probability sampling* peneliti harus memiliki data kependudukan yang memuat informasi mengenai seluruh individu dengan status kependudukan WNI. Salah satu kesalahan yang umum terjadi dalam membuat *sampling frame* adalah *coverage error*, yaitu situasi dimana tidak semua anggota populasi tercantum dalam *sampling frame*. *Coverage error* dapat menimbulkan **bias**, yaitu perbedaan sistematis antara kondisi sesungguhnya dari suatu populasi dengan kondisi yang diprediksikan melalui analisis pada kelompok sampel (Ruel et al., 2016). Dengan melakukan *probability sampling*, peneliti punya peluang lebih besar untuk mendapatkan kelompok sampel yang lebih bervariasi. Ingat, variasi (varians) adalah hal yang amat krusial dalam generalisasi. Semakin besar variabilitas kelompok sampel, maka semakin akurat prediksi atau kesimpulan yang ditarik oleh peneliti (Fowler, 2014).

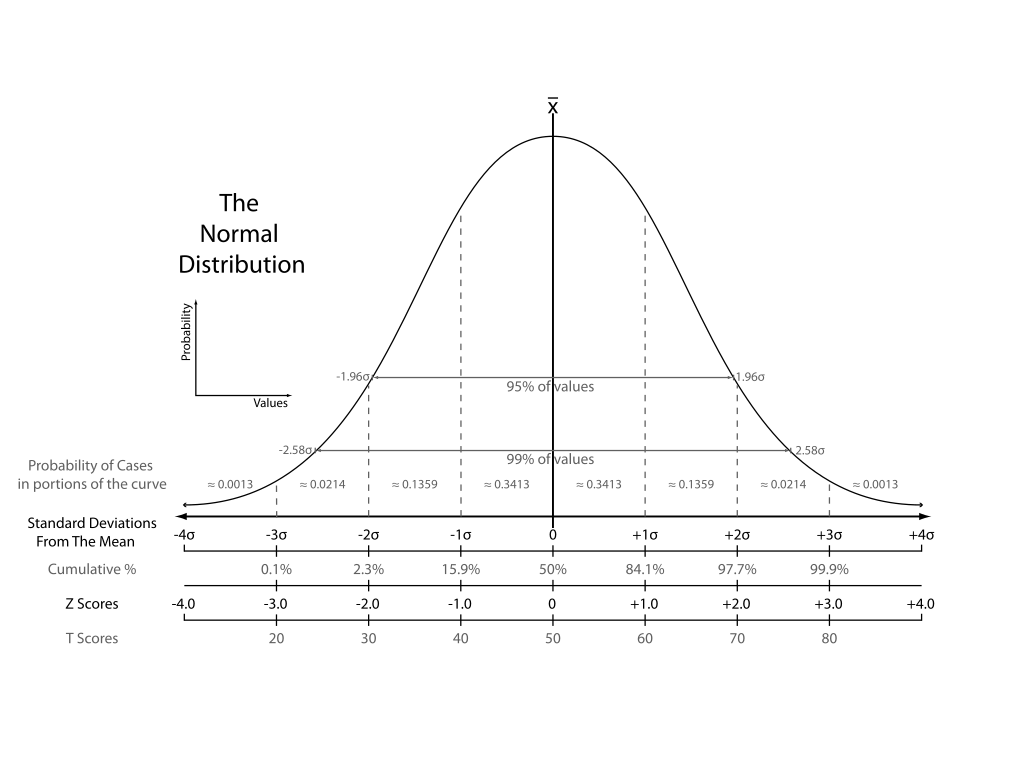
Ada beberapa strategi yang dapat digunakan untuk melakukan *probability sampling*, yaitu: **(a) *simple random sampling*** yang juga dikenal dengan *equal probability sample* yang mengasumsikan bahwa seluruh anggota populasi punya peluang yang sama persis untuk terpilih menjadi sampel melalui prosedur pengacakan yang sama. Seandainya anda ingin mengambil sampel dari populasi yang beranggotakan 1000 orang, maka peluang satu orang untuk terpilih menjadi sampel adalah setara, yaitu 1/1000. *Simple random sampling* dapat dilakukan dengan misalnya menggunakan *random number generator*. Google juga memiliki layanan ini dan dapat diakses dengan mengetikkan “*random number generator*” pada kolom pencarian. Meskipun kelihatannya ideal, *simple random sampling* justru menghasilkan variasi sampel yang amat kecil dan akibatnya kualitas sampel menjadi buruk, terutama apabila peneliti melibatkan populasi yang memiliki struktur di dalamnya.

Selanjutnya adalah **(b) *systematic random sampling***, yaitu teknik *sampling* yang pengambilan sampelnya mengikuti interval waktu atau jumlah tertentu. Misalnya, seorang petugas kendali mutu melakukan pemeriksaan kualitas produk dengan mengambil sampel sebanyak 1 produk untuk setiap 1000 produk yang diproduksi. Atau seorang peneliti Psikologi Politik yang mengambil sampel sebanyak 10 responden untuk setiap 100.000 penduduk. Namun perlu diingat, pengambilan sampel dengan interval tertentu menimbulkan masalah *periodicity* yang merupakan penyebab timbulnya bias karena akan menghasilkan kelompok sampel dengan variabilitas yang rendah, terutama apabila populasi memiliki struktur di dalamnya. Kedua teknik *probability sampling* yang telah dijelaskan sebelumnya tidak cocok diterapkan apabila peneliti menginginkan sampel yang representatif dari populasi yang memiliki struktur. Misalnya, apabila peneliti ingin melakukan studi pada Kelurahan A yang profil demografisnya beragam, maka peneliti menginginkan perwakilan sub-populasi (misalnya gender, tingkat pendidikan, dan tingkat penghasilan) dicakup dalam kelompok sampelnya. Dengan kedua teknik *sampling* yang dijelaskan sebelumnya, sangat mungkin sampel yang ditarik justru tidak proporsional karena mengabaikan struktur yang ada di populasi. Alternatifnya, peneliti dapat menggunakan: **(c) *cluster random sampling***, **(d) *stratified random sampling***, atau **(e) *multistage cluster sampling***. *Cluster sampling* berarti peneliti membuat *sampling frame* terpisah untuk masing-masing kategori non-strata pada populasi, misalnya sekolah, kelurahan, desa, RT, gender, dsb., lalu mengacak dan menentukan sampel di masing-masing kategori tersebut. Implikasinya, setiap kategori memiliki perwakilan dalam kelompok sampel. *Stratified sampling* adalah strategi yang sama, namun kategori yang digunakan untuk menentukan sampel adalah kategori dengan strata atau berjenjang, misalnya: tingkat pendidikan, tingkat penghasilan, dsb. Sedangkan yang terakhir, *multistage cluster sampling* atau *multilevel cluster sampling* menggunakan strategi yang lebih kompleks, yaitu mengambil sampel secara acak dari kategori yang berlapis. Pendekatan yang terakhir menghasilkan kualitas sampel yang terbaik (dengan variabilitas yang lebih besar), utamanya ketika peneliti tidak memiliki *sampling frame*.

Seorang peneliti Psikologi Politik tertarik untuk meneliti tingkat partisipasi politik pada pemilih pemula dengan kriteria: remaja yang berusia 18-23 tahun dan mahasiswa aktif pada jenjang S1. Agar mendapatkan sampel yang representatif, peneliti melakukan *multistage cluster sampling* dengan tahapan sebagai berikut: (a) menentukan secara acak sejumlah propinsi yang menjadi sampel dari 34 propinsi di Indonesia; (b) menentukan secara acak sejumlah perguruan tinggi yang menjadi sampel di masing-masing propinsi yang terpilih menjadi sampel pada tahap (a); (c) menentukan secara acak sejumlah program studi S1 pada masing-masing perguruan tinggi yang sudah terpilih menjadi sampel pada tahap (b); dan (d) menentukan secara acak sejumlah responden penelitian di masing-masing program studi S1 yang sudah terpilih pada tahap (c).

Problem *sampling* berikutnya berkaitan dengan jumlah sampel. Pertanyaan yang sering diajukan adalah berapa banyak jumlah sampel yang harus diambil oleh peneliti agar generalisasinya akurat? Meskipun kelihatannya sederhana, persoalan ini tidak mudah diatasi dan sangat bergantung pada tujuan survei. Apabila tujuan survei adalah membuat profil atau deskripsi, maka akurasi penarikan kesimpulan bergantung pada kadar ***sampling error*** (atau yang juga dikenal sebagai *margin error*). Namun ketika peneliti melakukan survei dengan tujuan menguji teori, hipotesis, atau prediksi, maka selain *sampling error*, peneliti harus memperhatikan *statistical power*, ukuran efek (*effect size*) yang ingin dideteksi, dan desain penelitian yang digunakan. Untuk topik yang terakhir, akan dibahas pada bagian 5.2.

*Sampling error* merupakan selisih antara rerata (*mean*) atau proporsi pada sampel dengan yang sebenarnya terjadi di populasi. Ingat, peneliti tak mungkin tahu apa yang terjadi di populasi sehingga yang dapat dilakukan adalah memperkirakan seberapa jauh menyimpang estimasi pada kelompok sampel dengan yang terjadi sebenarnya di populasi. Dalam statistik dikenal *central limit theorem* yang menyatakan apabila jumlah kelompok sampel (*n*) mendekati tak terhingga (), maka distribusi rerata dari kelompok sampel tersebut akan mengikuti **distribusi normal** (*gaussian distribution*), meskipun apabila distribusi data di populasi tidak mengikuti distribusi normal. Selanjutnya *the law of large numbers* menyebutkan apabila *n*=(), maka rerata sampel sama dengan rerata populasi.



*Figure* *2:*. Distribusi normal

**Distribusi normal** merupakan distribusi yang bentuk kurvanya secara visual mirip seperti lonceng, dimana mayoritas anggota kelompok sampel berada di tengah-tengah kurva (lihat gambar 2). Apabila distribusi sampel mengikuti distribusi normal, maka 68% dari jumlah total kelompok sampel akan jatuh pada rentang 1 standar deviasi masing-masing di kiri dan kanan rerata sampel, sedangkan 95% dari seluruh kelompok sampel akan jatuh pada rentang 1.96 standar deviasi masing-masing di kiri dan kanan rerata sampel. Karena tidak mungkin mengetahui standar deviasi populasi, maka yang dapat kita lakukan adalah menghitung ***standard error***, yaitu parameter statistik yang menunjukkan akurasi distribusi sampel dalam merepresentasikan populasi, dengan menggunakan standar deviasi kelompok sampel (Ruel et al., 2016).

Rumus *standard error of mean* (*SE*) adalah..

..dimana *S*y adalah **standar deviasi** pada kelompok sampel *y* dan *n* adalah jumlah sampel.

Apabila peneliti tertarik menghitung proporsi (persentase), maka rumus *standard error of proportion* (*SE*p) adalah..

..dimana *p*y adalah proporsi pada kelompok sampel *y*, sedangkan *n* adalah jumlah sampel.

Dengan mengetahui *standard error*, peneliti dapat menghitung *sampling error* (*margin error*) sekaligus rentang kepercayaan (*confidence interval*). Rentang kepercayaan adalah rentang di sekitar rerata atau proporsi sampel, dimana peneliti yakin bahwa rerata atau proporsi populasi jatuh di dalam rentang tersebut (Ruel et al., 2016). Semakin panjang rentang kepercayaan, estimasi menjadi kurang akurat, sedangkan rentang kepercayaan yang cenderung pendek menunjukkan bahwa estimasi pada kelompok sampel semakin dekat dengan nilai sesungguhnya di populasi.

Untuk memperoleh *sampling error*, kita tinggal mengalikan *standard error* dengan nilai kritis *z*, yaitu nilai yang berkaitan seberapa luas area kurva normal yang ingin diestimasi oleh peneliti. Nilai kritis *z* tergantung pada taraf kepercayaan () yang ditetapkan peneliti sebelum mengambil data. Apabila peneliti menetapkan =95%, maka nilai *z*=1.96, sehingga *sampling error* (*E*)..

Sedangkan untuk menghitung rentang kepercayaan (pada taraf kepercayaan 95%) adalah *E*.

**Ilustrasi 1**: Seorang peneliti Psikologi Politik melakukan survei pada 1000 responden di kota Lumajang dan menyimpulkan bahwa 32% responden menyatakan akan memilih ibu Paulina sebagai Walikota, lalu 34% responden menyatakan mendukung bapak Fernando Jose pada pemilihan umum bulan depan. Sedangkan sisanya tidak bersedia menjawab, tidak tahu, dan belum memutuskan. Berdasarkan informasi tersebut, maka *SE*p dapat diketahui sebesar 0.014.

Apabila peneliti tersebut menetapkan taraf kepercayaan () sebesar 95%, maka *sampling error* (*E*) adalah 1.96\*0.014=0.028 atau 2.8 persen. Oleh karena itu, dapat disimpulkan peneliti tersebut yakin (dengan peluang sebesar 95%) bahwa dukungan warga Lumajang pada ibu Paulina berkisar antara 29.2-34.8% (32% 2.8%) dan dukungan pada bapak Fernando Jose berkisar antara 31.2-36.2 persen (34% 2.8%).

Namun karena taraf kepercayaan yang ditetapkan sebesar 95%, maka masih ada peluang sebesar 5% bahwa persentase dukungan warga kota Lumajang pada masing-masing kandidat, jatuh di luar rentang kepercayaan tersebut. Karena rentang kepercayaan dukungan pada ibu Paulina dan bapak Fernano Jose tumpang tindih, maka meskipun pada kelompok sampel bapak Fernando Jose mendapatkan suara terbanyak, belum tentu ia memenangkan pemilihan umum.

**Ilustrasi 2**: Seorang peneliti Psikologi Politik melakukan survei pada 100 responden di kota Magetan untuk mengetahui berapa kali responden melihat iklan calon legislatif dalam sehari. Apabila responden penelitian rata-rata melihat iklan calon legislatif sebanyak 12 kali dalam sehari dengan standar deviasi sebesar 8 kali, maka *SE*=0.8. Apabila ditetapkan sebesar 95%, maka *E* sebesar 1.96\*0.8=1.56 kali (dibulatkan ke atas menjadi 2 kali).

Dari informasi tersebut, umumnya (dengan peluang sebesar 95%) warga kota Magetan rata-rata melihat iklan calon legislatif antara 10-14 (10 2) kali dalam sehari. Namun karena ditetapkan sebesar 95%, maka masih ada peluang sebesar 5% rerata intensitas warga Magetan melihat iklan calon legislatif dalam sehari, jatuh di luar rentang tersebut.

Apabila memperhatikan kembali formula *standard error* dan *sampling error* kita dapat menyimpulkan bahwa estimasi survei akan semakin akurat ketika sampel memiliki variasi yang rendah (ditunjukkan oleh standar deviasi yang kecil) dan ketika peneliti mengambil sampel dalam jumlah besar. Yang menarik, ukuran populasi tidak berkaitan dengan akurasi sehingga bertentangan dengan yang banyak dipercayai oleh orang awam, *n* = 1000 akan menghasilkan tingkat akurasi yang sama besarnya berapapun ukuran populasi yang diwakilinya.

Pada teknik *probability sampling*, peneliti juga harus mempertimbangkan taraf respon (*response rate*). Seandainya peneliti mengundang 1000 orang yang terpilih menjadi sampel dan hanya 100 orang yang bersedia dan menyelesaikan partisipasi, artinya taraf respon hanya sekitar 10% saja. Padahal apabila taraf respon kurang dari 70 persen, maka sangat besar kemungkinannya bahwa banyaknya partisipan yang tidak merespon tidak terjadi secara kebetulan (*random*) sehingga sangat mungkin terjadi bias (Ruel et al., 2016). *Nonresponse error* merupakan sumber kesalahan yang langsung berdampak pada kualitas temuan penelitian survei sehingga apabila taraf respon rendah, maka kualitas data yang didapatkan menjadi kurang baik.

Menentukan sampel dengan teknik *probability sampling* adalah strategi yang ideal, namun yang lebih sering terjadi, peneliti menghadapi kondisi dimana *probability sampling* tidak dapat dilakukan. Kendala yang paling umum adalah peneliti tidak memiliki informasi yang cukup untuk membuat *sampling frame* sehingga peluang terpilihnya satu unit sampel tidak dapat diketahui. Bisa juga kondisi ini terjadi ketika penelitian survei melibatkan populasi dengan kriteria yang sangat spesifik sehingga sulit diakses. Dalam kondisi ini, peneliti dapat menggunakan ***non-probability sampling*** dimana sampel ditarik tanpa proses pengacakan. Implikasinya memang tidak menyenangkan - bias dan *sampling error* tidak dapat diestimasi sehingga generalisasi sulit sekali dilakukan (Fowler, 2014).

Beberapa teknik *non-probability sampling* yang umumnya dilakukan (Ruel et al., 2016) adalah; **(a) *convenience sampling***, yaitu pemilihan sampel berdasarkan kemudahan dan ketersediaan akses. Contohnya ketika peneliti merekrut keluarga, teman, kolega, dan kenalan lainnya sebagai responden survei. Stategi ini tentu berisiko membiarkan subjektivitas peneliti mengkontaminasi keputusan pemilihan sampel (*selection bias*). Selanjutnya adalah **(b) *quota sampling***, yaitu *convenience sampling* yang ditambah batasan jumlah (kuota) pada kategori demografis tertentu (misalnya, 50% laki-laki dan sisanya perempuan).

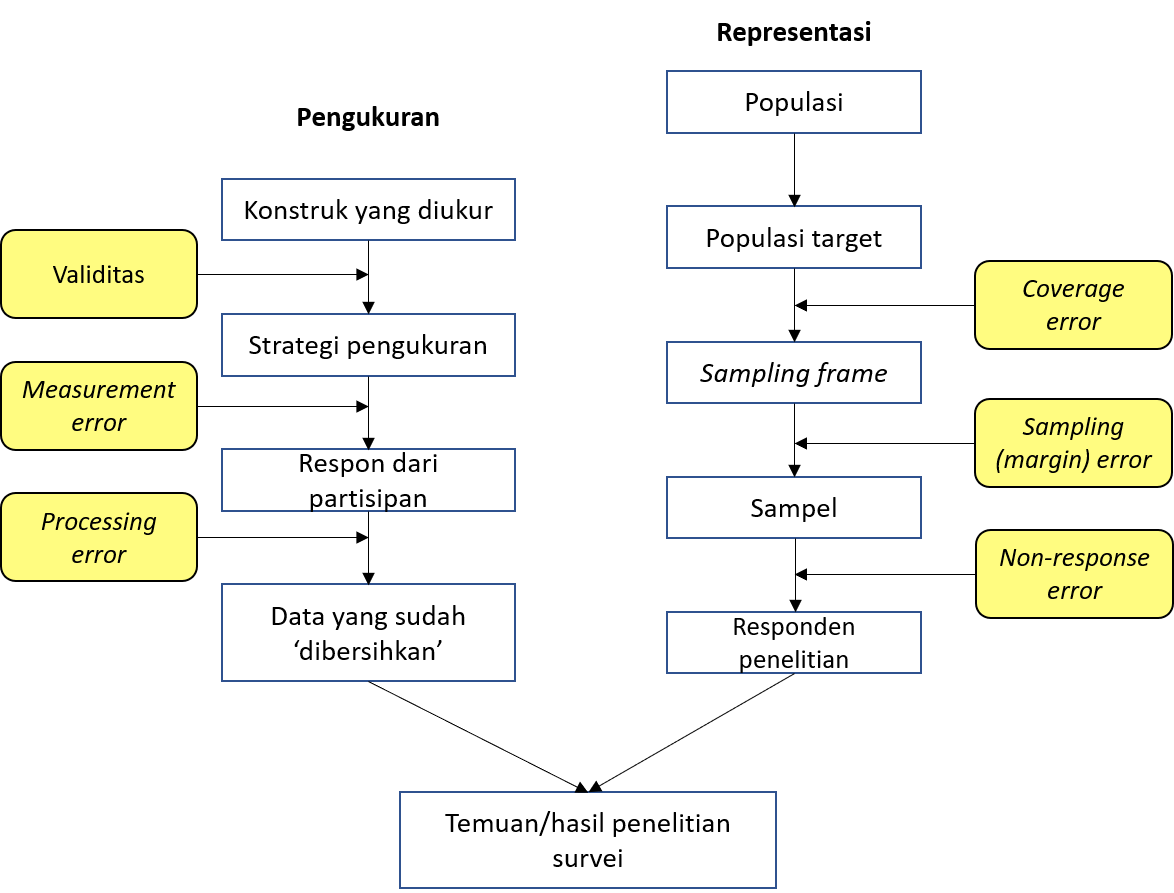
Berikutnya adalah **(c) *purposive* atau *judgemental sampling***, yang merupakan salah satu bentuk *convenience sampling*. Teknik ini digunakan ketika sampel dipilih berdasarkan pengetahuan mereka yang spesifik, yang berkaitan dengan topik penelitian. Terminologi lain untuk teknik ini adalah *key informant interview*, yaitu ketika hanya informan yang dianggap menguasi isu yang sedang diteliti, yang dicakup sebagai sampel. Teknik *purposive sampling* cocok digunakan ketika meneliti komunitas marjinal karena umumnya *sampling frame* pada komunitas marjinal (misalnya, anak jalanan, pekerja sosial, aktivis LSM, dsb.) tidak tersedia dan tidak terkonsentrasi pada satu area tertentu.

Yang terakhir, **(d) *snowball sampling*** atau yang juga dikenal sebagai *chain referral* atau *chain sampling*, yaitu strategi rekrutmen responden dengan cara meminta responden penelitian untuk merekomendasikan atau merujuk calon responden lainnya yang sesuai dengan kriteria yang ditetapkan oleh peneliti. Teknik ini sangat cocok apabila peneliti menghadapi populasi yang sulit diakses, lebih kecil cakupannya dari populasi umum, dan umumnya berada di lokasi yang saling berjauhan.

*4.1.3 Menilai kualitas temuan penelitian survei*

Di banyak laporan penelitian survei, *sampling error* (atau *margin error*) dilaporkan sedemikian rupa sehingga menimbulkan kesan bahwa seolah-olah *sampling error* adalah satu-satunya indikator yang mencerminkan kualitas survei secara keseluruhan. Padahal faktanya, sumber kesalahan (*source of error*) lainnya justru lebih penting untuk dikendalikan, terutama pada survei yang melibatkan jumlah sampel atau populasi target yang besar. Selain itu, tidak jarang peneliti melaporkan *sampling error*, padahal peneliti tidak mempertimbangkan struktur populasi ketika proses pengacakan, atau bahkan ketika ia tidak melakukan *probability sampling* sama sekali. Padahal tidak mungkin *sampling error* dapat dihitung ketika peneliti tidak melakukan *probability sampling* sehingga dalam situasi seperti ini, memberikan informasi mengenai *sampling error* tentu akan memberikan impresi yang keliru mengenai proses penelitian.

Banyak peneliti survei sepakat perlu adanya kerangka kerja yang lebih luas untuk menilai kualitas survei sehingga paradigma *total survey error* (TSE) digagas sebagai kerangka kerja yang dapat membantu peneliti memperkirakan kualitas temuan penelitian survei. TSE pada dasarnya adalah **akumulasi seluruh sumber kesalahan (yang menyebabkan ketidaktepatan estimasi) yang melekat pada desain, pengumpulan, pemrosesan, dan analisis data penelitian survei** (Biemer, 2010; Groves & Lyberg, 2010). Dalam konteks ini, *error* merujuk pada penyimpangan (atau selisih) antara estimasi pada kelompok sampel dengan nilai parameter yang sesungguhnya di populasi. TSE mencakup beberapa sumber kesalahan yang semuanya memiliki dampak langsung dalam menentukan kualitas temuan penelitian survei.



*Figure* *3:*. *Total Survey Error*. Diadopsi dari Groves dan Lyberg (2010).

TSE pada dasarnya mencakup semua kemungkinan kesalahan yang muncul dari proses penelitian survei yang tidak hanya mencakup *sampling error* saja. TSE mencakup semua kemungkinan kesalahan yang terjadi pada dua proses utama dalam survei, yaitu penarikan sampel dan pengukuran konstruk yang diteliti (Groves & Lyberg, 2010). Beberapa diantaranya (validitas, *measurement error*/reliabilitas, *coverage error*, dan *nonresponse error*) telah dijelaskan di bagian sebelumnya. Sedangkan *processing error* merupakan kesalahan yang mungkin terjadi ketika proses *input* data (apabila menggunakan survei dengan administrasi *paper-based*), pembersihan (*data cleaning*), pembobotan (*weighting*), pengorganisasian dan transformasi data (*data wrangling*).

Meskipun TSE telah mencakup lebih banyak aspek yang mempengaruhi akurasi temuan survei, banyak peneliti belum puas dan menyebutnya terlalu sederhana karena mengabaikan aspek kualitas dari kacamata pengguna survei. Apalagi pengguna survei biasanya tidak terlalu memperhatikan, bahkan menerima begitu saja, aspek akurasi ini (Biemer, 2010). Dalam menilai kegunaan dan kualitas temuan survei pengguna lebih memprioritaskan ketepatan waktu, aksesibilitas, kegunaan data, dan desain instrumen yang relevan dengan pertanyaan penelitian yang akan dijawab (Biemer, 2010). Oleh karena itu, dapat disimpulkan bahwa kualitas survei merupakan konsep yang kompleks, multidimensional, dan lebih komperhensif melampaui cakupan TSE.

Dengan mengintegrasikan standar kualitas yang ditetapkan oleh berbagai Biro Statistik milik pemerintah di berbagai negara dan kawasan, seperti Eurostat, Australian Bureau of Statistics, Statistics Canada, Stats NZ (Selandia Baru), dan U.S. Census Bureau, Biemer (2010) menggagas kerangka kerja yang lebih komperhensif dalam menilai kualitas survei yang disebut ***total survey quality framework*** (SQF). SQF dapat menjadi daftar tilik (*checklist*) untuk menilai kualitas survei secara keseluruhan dan mencakup sembilan dimensi (lihat tabel 1).

Tabel 1. Dimensi *Survey Quality Framework* (Biemer, 2010)

|  |  |
| --- | --- |
| Dimensi | Deskripsi |
| Akurasi | *Total survey error* diminimalisasi |
| Kredibilitas | Temuan survei dianggap terpercaya atau kredibel oleh komunitas profesional |
| Komparabilitas | Perbandingan temuan survei pada kategori demografis, spasial, dan temporal yang berbeda tepat dan meyakinkan (*robust*) |
| Kegunaan/Kemudahan interpretasi data | Dokumentasi dan metadata disimpan dengan baik dan mudah digunakan |
| Relevansi | Data/temuan hasil survei sesuai dengan kebutuhan pengguna |
| Aksesibilitas | Akses pada data mentah tersedia dan mudah diakses |
| Ketepatan waktu | Proses survei diselesaikan tepat pada waktunya |
| Kelengkapan | Data sudah cukup kaya sehingga analisis data dapat menjawab pertanyaan penelitian, tanpa harus membebani partisipan secara berlebihan |
| Koherensi | Apabila temuan dicek silang dengan data lain dari sumber yang berbeda, akan menghasilkan kesimpulan yang tidak jauh berbeda |

**4.1.4 Teknik pengumpulan data survei**

Memilih metode administrasi survei merupakan keputusan penting yang harus diambil oleh peneliti. Survei dapat dilakukan dengan wawancara terstruktur, yaitu dilakukan dengan bertanya langsung pada responden lalu merekam jawabannya, atau instrumen dapat diisi secara mandiri oleh responden (*self-report*). Ketika diputuskan bahwa survei diadministrasi dengan wawancara, peneliti dapat mewawancarai responden melalui telepon (*telephone-based survey*), *interactive voice response* (IVR), atau langsung dengan tatap muka. Apabila instrumen diisi secara mandiri oleh partisipan, survei dapat menggunakan cara klasik, yaitu menggunakan kertas (*paper-based*). Alternatif yang saat ini populer digunakan adalah menggunakan *platform* penyedia layanan kuesioner daring, seperti [Google Forms](https://www.google.com/forms/about/), [LimeSurvey](https://www.limesurvey.org/), [Qualtrics](https://www.qualtrics.com/), [SosSci Survey](https://www.soscisurvey.de/), dsb. Kuesioner dapat diisi secara klasikal, dimana partisipan dikumpulkan dalam satu ruangan lalu mengisi kuesioner bersama-sama, atau dapat juga diisi secara mandiri dengan perangkat masing-masing.

Peneliti juga punya banyak pilihan strategi dalam merekrut responden. Bisa dengan mengirim surat elektronik, mengeposkan kuesioner ke alamat responden, merekrut melalui *platform crowdworkers* seperti Amazon Mechanical Turk (MTurk), Clickworker, atau Prolific. Peneliti juga dapat memasang iklan atau undangan berpartisipasi di media sosial atau menyebarkannya secara berantai melalui aplikasi *instant messaging*, seperti WhatsApp, LINE, atau Telegram. Yang perlu diperhatikan, merekrut responden melalui *platform crowdworkers* atau melalui pesan berantai dan pemasangan iklan di media sosial atau aplikasi *instant messaging* implikasinya peneliti tidak melakukan proses *sampling* karena keputusan berpartisipasi sangat bergantung pada kesediaan responden. Artinya, responden yang berpartisipasi adalah orang adalah yang secara sukarela bersedia berpartisipasi, bukan yang dipilih melalui proses penentuan sampel yang dilakukan oleh peneliti. Oleh karena itu, merekrut dengan strategi ini berisiko menimbulkan *self-selection bias* (Bethlehem, 2010).

Dalam memilih teknik pengumpulan data, metode administrasi, dan strategi perekrutan responden yang tepat, peneliti perlu memperhatikan beberapa hal yang berkaitan dengan tujuan dan desain penelitiannya (Fowler, 2014). Faktor pertama yang perlu dipertimbangkan adalah **strategi penarikan *sampling***. Strategi yang digunakan peneliti untuk menarik sampel akan berimplikasi pada cara pengambilan data. Apabila peneliti hanya memiliki *sampling frame* yang berisi nama anggota populasi tanpa alamat pos, nomor telpon, alamat surat elektronik, atau nomor ponsel, maka mengumpulkan data dengan model korespondensi tidak mungkin dilakukan. Beberapa peneliti memilih langsung mendatangi tempat tertentu dimana responden dapat ditemui (misalnya, tempat kerja, sekolah, dsb.) sehingga dalam konteks ini banyak pilihan teknik pengambilan data yang bisa dilakukan, baik dengan wawancara langsung, menggunakan kuesioner cetak, atau menggunakan kuesioner dengan alat bantu seperti tablet, komputer *desktop*, atau ponsel - semuanya dapat dilakukan.

Faktor yang kedua adalah **karakteristik populasi target**. Model *self-report* lebih cocok diterapkan pada responden dengan kemampuan literasi yang memadai sehingga pada populasi dengan tingkat pendidikan yang cenderung rendah, wawancara terstruktur akan menghasilkan data yang lebih baik. Apabila peneliti memilih untuk mengadministrasikan kuesioner dengan menggunakan perangkat tertentu dan membutuhkan internet, maka peneliti berisiko mengecualikan responden yang tidak memiliki dan/atau tidak mampu mengoperasikan perangkat, serta yang tidak memiliki akses internet. Menggunakan model *self-report* juga berpotensi meningkatkan jumlah responden yang tidak menyelesaikan pengisian kuesioner dan semakin banyak jumlah pertanyaan yang tidak dijawab (*missing*). Apabila penyelesaian pengisian kuesioner dan kelengkapan respon adalah prioritas utama peneliti, maka administrasi kuesioner dengan wawancara akan menghasilkan data dan taraf respon yang lebih memuaskan.

Faktor berikutnya adalah **bentuk dan substansi kuesioner**. Apabila kuesioner banyak berisi pertanyaan terbuka (*open-ended question*), maka mengadministrasinya dengan model *self-report* akan memperbesar kemungkinan kuesioner tidak seluruhnya diisi karena merespon pertanyaan terbuka membutuhkan beban kognitif yang sangat besar. Kalaupun pertanyaan terbuka diisi oleh responden, seringkali respon tersebut tidak terlalu berguna. Untuk kuesioner dengan pilihan ganda yang melibatkan banyak opsi jawaban, wawancara terstruktur ataupun via telpon akan mempersulit pewawancara dan responden sehingga *self-report* lebih cocok digunakan. Ketika aitem dalam kuesioner menggunakan stimulus atau petunjuk visual, maka wawancara terstruktur lewat telepon juga tidak bisa dilakukan.

Selain bentuk aitem, substansi aitem dalam kuesioner juga perlu menjadi pertimbangan. Banyak ahli sepakat apabila substansi kuesioner menanyakan hal-hal yang dianggap sensitif, maka mewawancarai responden, baik secara langsung maupun melalui telepon, akan mendorong responden untuk memberikan jawaban yang diinginkan secara sosial (*socially desirable*) bukan jawaban yang sebenarnya. Apabila peneliti membutuhkan respon yang sangat detail mengenai aktivitas partisipan, misalnya berapa kali partisipan melihat baliho iklan calon anggota legislatif, atau berapa lama dalam sehari waktu yang dihabiskan untuk membaca berita politik, maka peneliti perlu meminimalisasi beban responden untuk mengingat kembali aktivitas yang mereka lakukan. Strategi terbaik yang dapat diterapkan meminta responden mengisi semacam diari yang dapat disimpan untuk merekam aktivitas mereka sehari-hari. Atau peneliti dapat menggunakan teknologi yang dapat memungkinkan responden melaporkan aktivitas mereka secara periodik via *short messaging service* (SMS), surat elektronik, atau dapat menggunakan aplikasi khusus yang didesain untuk kepentingan pengambilan data.

**4.1.5 Survei opini publik, hitung cepat (*quick count*), dan *exit poll***

Menjelang pemilihan umum, aktivitas survei opini publik yang diselenggarakan oleh lembaga survei non-pemerintah marak menghiasi pemberitaan. Sejak pertama kali diperkenalkan oleh ilmuwan Psikologi, George Gallup, yang melakukan survei opini publik menjelang pemilihan Presiden Amerika Serikat pada tahun 1933, survei opini publik merupakan bagian penting dari proses demokrasi di beberapa negara (Donsbach, 2015). Di Indonesia, survei opini publik marak dilakukan pasca pemerintahan Orde Baru dan mulai menjadi bagian penting dari politik elektoral di Indonesia sejak pemilihan umum tahun 2004. Sejak era Orde Baru berakhir, masyarakat nampaknya sudah memiliki aspirasi yang konkrit mengenai kualitas figur yang mereka harapkan untuk menempati jabatan publik. Kualitas ini lebih mengarah pada kinerja daripada ikatan primordial sehingga politisi mulai tertarik untuk mendapatkan lebih banyak informasi mengenai calon pemilih mereka. Survei opini publik yang dilakukan oleh lembaga survei dapat mengakomodasi kebutuhan ini (Mietzner, 2009).

Tidak hanya survei opini publik yang lazimnya dilakukan sebelum atau pada masa kampanye, lembaga survei juga melakukan hitung cepat (*quick count*) dan beberapa diantaranya melakukan *exit poll*. Berbeda dengan survei opini publik yang menggunakan individu pemilih sebagai unit analisis, hitung cepat menggunakan tempat pemungutan suara (TPS) sebagai sampel untuk memprediksi hasil pemilihan umum. Hasil penghitungan suara di TPS sampel kemudian diakumulasi dan dijadikan informasi sementara yang dapat diakses publik mengenai hasil pemilihan umum, sembari menunggu hasil perhitungan dan verifikasi suara yang dilakukan Komisi Pemilihan Umum yang biasanya membutuhkan waktu kurang lebih sebulan setelah pemungutan suara. *Exit poll* sedikit berbeda karena data didapatkan dengan mewawancarai pemilih segera setelah ia keluar dari TPS. Dari segi kecepatan, *exit poll* memungkinkan lembaga survei menyimpulkan temuan lebih cepat, setidaknya tidak jauh setelah waktu pemungutan suara berakhir, sedangkan hitung cepat baru dapat dikumpulkan ketika penghitungan suara selesai dilakukan di TPS. Namun dari segi akurasi, hitung cepat biasanya lebih akurat karena menggunakan data riil dari penghitungan suara di tingkat TPS.

Meskipun kerapkali dipertanyakan kredibilitasnya, seperti yang diuraikan penulis sebagai pembuka bab ini, mayoritas lembaga survei di Indonesia menyediakan informasi yang substansial mengenai perilaku pemilih. Dalam pemilihan Presiden tahun 2019 misalnya, sebagian besar temuan lembaga survei menyebutkan Joko Widodo (Jokowi)-Ma’ruf Amin mendapatkan lebih banyak dukungan, yaitu sekitar mendekati sampai sedikit diatas 50 persen, sedangkan kompetitornya hanya mendapat sekitar 31-37 persen (Soderborg, 2019). Temuan ini sejalan dengan hasil pemilihan umum 2019, dimana Jokowi-Ma’ruf Amin meraup suara hingga 55.5 persen. Hasil hitung cepat sebagian besar lembaga survei juga terbukti amat akurat dengan rata-rata *sampling error* hanya kurang dari 1 persen (Intan, 2019).

Yang menarik, lembaga-lembaga survei di Indonesia memiliki rekam jejak yang kurang baik dalam memprediksikan dukungan pemilih pada partai-partai Islam. Pada pemilihan legislatif 2014 dan 2019 misalnya, suara yang diperoleh Partai Keadilan dan Sejahtera (PKS), Partai Amanat Nasional (PAN), dan Partai Persatuan Pembangunan (PPP) secara konsisten lebih tinggi daripada yang diprediksikan oleh kebanyakan lembaga survei (Fealy, 2014; Soderborg, 2019). Adanya kesenjangan antara survei opini publik mengenai dukungan pada partai-partai Islam dengan hasil pemilihan umum memantik pertanyaan-pertanyaan menarik; apakah responden sengaja ‘menyembunyikan’ dukungannya pada partai Islam? Atau apakah ada kemungkinan desain kuesioner yang digunakan mengandung bias yang merugikan pendukung partai Islam?

Soderborg (2019) menemukan bukti pemilih PPP dan PKS lebih mungkin “menyembunyikan” pilihan mereka ketika diwawancarai oleh lembaga survei, dibandingkan dengan pemilih partai lainnya. Mengingat dukungan pada partai Islam juga dapat dijelaskan oleh lokasi geografis pemilih karena pemilih partai Islam biasanya terkonsentrasi pada lokasi-lokasi tertentu, maka desain pengambilan sampel yang diterapkan lembaga survei dapat juga menjadi penyebab inakurasi ini. Soderborg (2019) juga menemukan bukti yang menguatkan dugaan ini, yaitu ada korelasi yang jelas antara jumlah sampel yang diambil lembaga survei dengan dukungan pada partai-partai Islam. Artinya, survei yang menggunakan jumlah sampel lebih banyak akan melaporkan tingkat keterpilihan partai-partai Islam yang lebih tinggi daripada survei dengan jumlah responden yang lebih sedikit. Lebih lanjut, ada juga kemungkinan partai-partai Islam diuntungkan oleh proses kampanye yang lebih intensif di hari-hari akhir menjelang pemungutan suara.

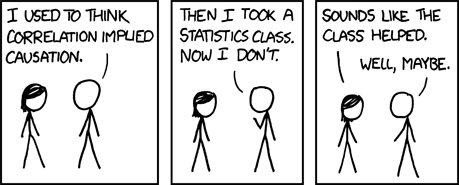
Pertanyaan mengenai kredibilitas lembaga survei bukan hal baru dan sebenarnya dapat diperiksa dengan cukup mudah dengan standar yang jelas, yaitu dengan memperhatikan *survey quality framework* (SQF) (Groves & Lyberg, 2010). Namun yang kerapkali terjadi, dalam mengkomunikasikan temuanya, lembaga survei terlalu berfokus pada *sampling error* dan tidak banyak informasi yang tersedia untuk mengevaluasi apakah lembaga survei melakukan pekerjaannya dengan baik. Laporan yang dirilis lembaga survei biasanya hanya berupa rangkuman eksekutif dengan informasi yang sangat minimal sehingga sulit menakar akurasi temuan survei yang mereka lakukan, lebih-lebih apabila memperhatikan aspek-aspek *total survey error*. Laporan yang dirilis juga seringkali hanya mencantumkan persentase atau proporsi absolut, tidak sekaligus dengan rentang kepercayaanya. Hal ini konsisten dengan yang dituliskan Biemer (2010), yaitu meskipun peneliti di lembaga akademik dan biro statistik milik pemerintah sudah mulai menetapkan standar kualitas yang ketat dengan mengadopsi SQF, praktik yang sama tidak ditemui pada lembaga survei non-pemerintah.

Dalam mengkomunikasikan temuan penelitian survei, utamanya untuk kepentingan akademik, peneliti sangat disarankan untuk selalu melaporkan rentang kepercayaan (*confidence interval*) untuk memberikan informasi kepada pembaca mengenai akurasi estimasi yang dilakukan oleh peneliti. Sebagian besar *outlet* publikasi ilmiah terkemuka sudah mensyaratkan pelaporan rentang kepercayaan, utamanya jurnal-jurnal yang diterbitkan oleh American Psychological Association (APA) (Appelbaum et al., 2018). Misalnya, apabila berdasarkan hasil survei kandidat A memperoleh 64% suara dengan *sampling error* sebesar 6 persen, maka seharusnya dalam laporan disebutkan, “…berdasarkan hasil survei, kandidat A mendapatkan dukungan terbanyak, yaitu sekitar 64% (95% CI 59-70%) responden penelitian..”

Kurangnya transparansi dan kontrol publik merupakan kritik yang paling sering dilontarkan pada lembaga survei. Laporan lengkap, data mentah, dokumentasi, dan metadata dari survei opini publik yang mereka umumkan hasilnya juga tidak bisa diakses terbuka sehingga sulit untuk menakar kredibilitas temuannya. Meskipun beberapa lembaga survei yang dipandang profesional sudah memulai tradisi baik dengan bersedia memaparkan data hitung cepat yang dikumpulkan pada pemilihan umum serentak tahun 2019 yang lalu (Intan, 2019), informasi yang diberikan sangat terbatas sehingga sulit menakar apakah temuan tersebut dapat direka-ulang (*reproducibility*). Data hitung cepat yang dipaparkan Cyrus Network dan Center for Strategic and International Studies (CSIS) di momen tersebut, misalnya, disajikan dalam bentuk statis dalam format .html dan nyaris tidak memiliki kegunaan sama sekali untuk diperiksa, apalagi disimpulkan akurasi estimasinya. Informasi krusial lainnya seperti desain *sampling* juga tidak ditemukan.

## 4.2 Eksperimental

**4.2.1 Kausalitas versus asosiasi**



*Figure* *4:*. Korelasi dan kausalitas. Sumber: <https://xkcd.com>

**4.2.2 Ragam desain eksperimen (*within* dan *between-group design*)**

**4.2.3 Validitas internal dan eksternal penelitian eksperimen**

**4.2.4 Pendekatan survei eksperimen**

## 4.3 Pemrosesan teks (*natural language processing*) dan metode implisit

## 4.4 Penelitian meta (tinjauan sistematis dan meta-analisis)

**4.4.1 Kelebihan dan fungsi penelitian kumulatif**

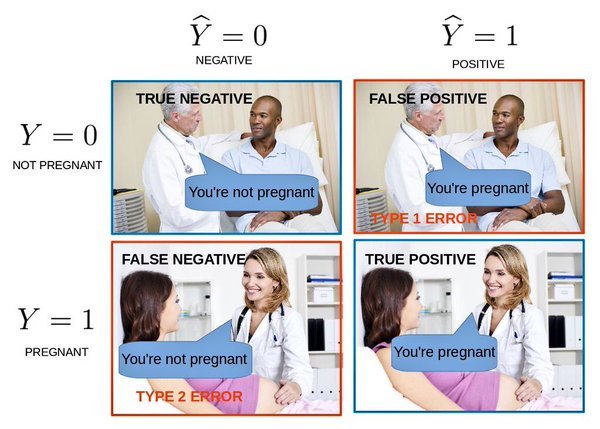
**4.4.2 Tren penelitian meta dalam perkembangan Psikologi Politik**

## 4.5 Penyimpulan karakter tokoh politik dengan pendekatan *at a distance*

## 4.6 Pendekatan interpretif (kualitatif)

## 5. Pengujian hipotesis

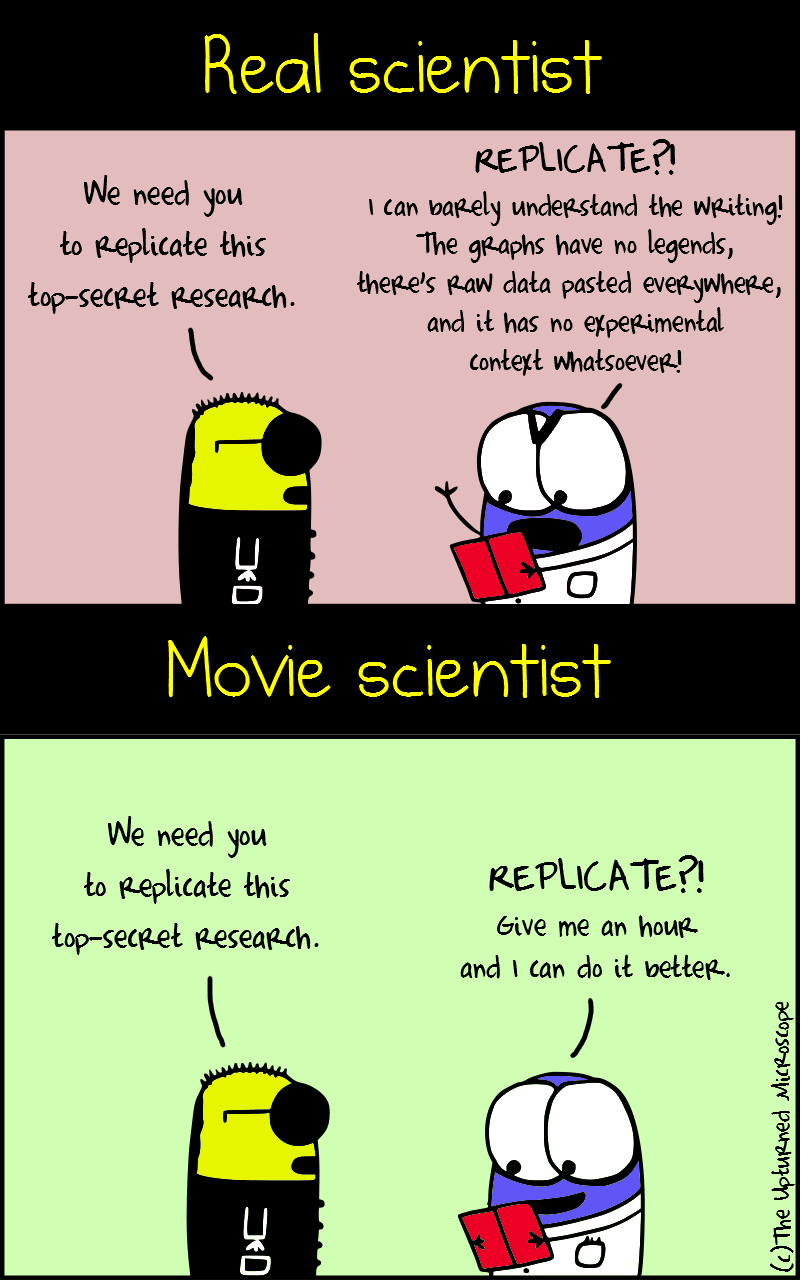
## 5.1 Kesalahan penarikan kesimpulan dalam paradigma *null hypothesis significant testing*



*Figure* *5:*. Kesalahan dalam Penarikan Kesimpulan (Inferensi). Sumber: <https://dzone.com/articles/understanding-the-confusion-matrix>

## 5.2 Nilai p, *statistical power*, ukuran efek (*effect size*), dan rentang kepercayaan (*confidence interval*)

## 6. Revolusi kredibilitas: Coba-ulang (*reproducibility*) dan reka-ulang (*replicability*) dalam penelitian Psikologi



*Figure* *6:*. Reka-ulang Data Penelitian. Sumber: <https://theupturnedmicroscope.com/>

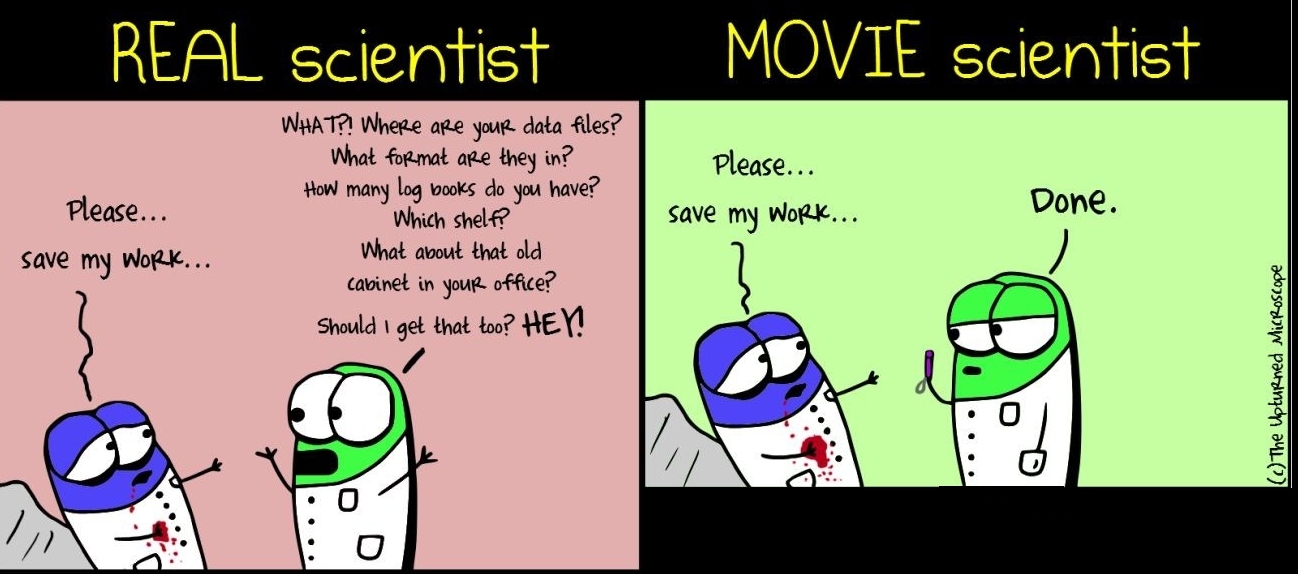
## 6.1 Krisis kredibilitas, bias publikasi, dan praktik meneliti yang meragukan (*questionable research practices*)

## 6.2 Tantangan dan peluang bagi peneliti Indonesia

## 6.3 Petunjuk praktis agar penelitian dapat direka-ulang

## 6.3.1 Pra-registrasi

## 6.3.2 Manajemen data penelitian berbasis kendali versi (*version control*)



*Figure* *7:*. Reka-ulang Data Penelitian. Sumber: <https://theupturnedmicroscope.com/>

## 7. Isu-isu etika dalam penelitian Psikologi Politik

## 7.1 Anonimitas dan kerahasiaan

## 7.2 Persetujuan atas tindakan (*informed consent*)

## 7.3 Transparansi dan risiko pengungkapan (*disclosure risk*)

# References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. <https://doi.org/10.1037/amp0000191>

Bethlehem, J. (2010). Selection Bias in Web Surveys: Selection Bias in Web Surveys. *International Statistical Review*, *78*(2), 161–188. <https://doi.org/10.1111/j.1751-5823.2010.00112.x>

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. <https://doi.org/10.1093/poq/nfq058>

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, *23*(2), 532–544. <https://doi.org/10.1037/a0022402>

Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, *19*(4), 651–682. <https://doi.org/10.1177/1094428116656239>

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, *52*(4), 281–302.

Dienes, Z. (2008). *Understanding psychology as a science*. London: Palgrave Macmillan.

Donsbach, W. (2015). Public Opinion Polls. In G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication*. New York: Wiley-Blackwell.

Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In *Advances in experimental social psychology* (Vol. 33, pp. 41–113). <https://doi.org/10.1016/S0065-2601(01)80004-6>

Dumenci, L. (2000). Multitrait-Multimethod Analysis. In H. E. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 583–611). San Diego: Elsevier. <https://doi.org/10.1016/B978-0-12-691360-6.X5000-9>

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. <https://doi.org/10.1111/bjop.12046>

Fealy, G. (2014). Resurgent political Islam, or astute Islamic parties? *New Mandala*.

Fowler, F. J. (2014). *Survey research methods* (Fifth edition). Los Angeles: SAGE.

Fried, E. I., & Flake, J. K. (2018). Measurement Matters. *APS Observer*, *31*(3).

Groves, R. M., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, *74*(5), 849–879. <https://doi.org/10.1093/poq/nfq065>

Hunsley, J., & Meyer, G. J. (2003). The Incremental Validity of Psychological Testing and Assessment: Conceptual, Methodological, and Statistical Issues. *Psychological Assessment*, *15*(4), 446–455. <https://doi.org/10.1037/1040-3590.15.4.446>

Intan, G. (2019). Dituduh Tukang Bohong, Lembaga Survei Ungkap Data dan Metodologi Quick Count. *VOA Indonesia*. https://www.voaindonesia.com/a/dituduh-tukang-bohong-lembaga-survei-ungkap-data-dan-metodologi-quick-count/4885243.html.

Mair, P. (2018). *Modern Psychometrics with R*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-93177-7>

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*(4), 523–547.

McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods*, *23*(3), 412–433. <https://doi.org/10.1037/met0000144>

Messick, S. (1995). Validity of Psychological Assessment. *American Psychologist*, 9.

Mietzner, M. (2009). Political opinion polling in post-authoritarian Indonesia: Catalyst or obstacle to democratic consolidation? *Bijdragen Tot de Taal-, Land- En Volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, *165*(1), 95–126. <https://doi.org/10.1163/22134379-90003644>

Morling, B. (2018). *Research Methods in Psychology: Evaluating a World of Information* (3rd ed.). New York, NY: W. W. Norton & Company, Inc.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>

Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, *6*(MAR), 1–11. <https://doi.org/10.3389/fpsyg.2015.00223>

Peters, G. J. Y. (2018). *The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach’s alpha and the route towards more comprehensive assessment of scale quality* (Preprint). PsyArXiv. <https://doi.org/10.31234/osf.io/h47fv>

Revelle, W., & Zinbarg, R. E. (2008). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145. <https://doi.org/10.1007/s11336-008-9102-z>

Ruel, E. E., Wagner, W. E., & Gillespie, B. J. (2016). *The practice of survey research: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Sijtsma, K. (2008). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha. *Psychometrika*, *74*(1), 107. <https://doi.org/10.1007/s11336-008-9101-0>

Skrondal, A., & Rabe-Hesketh, S. (2007). Latent Variable Modelling: A Survey. *Scandinavian Journal of Statistics*, *34*(4), 712–745. <https://doi.org/10.1111/j.1467-9469.2007.00573.x>

Soderborg, S. (2019). Indonesia: How the polls are performing. *New Mandala*.

Thornton, S. (2019). Karl Popper. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two Quantitative Approaches for Estimating Content Validity. *Western Journal of Nursing Research*, *25*(5), 508–518. <https://doi.org/10.1177/0193945903252998>

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. <https://doi.org/10.1037/amp0000191>

Bethlehem, J. (2010). Selection Bias in Web Surveys: Selection Bias in Web Surveys. *International Statistical Review*, *78*(2), 161–188. <https://doi.org/10.1111/j.1751-5823.2010.00112.x>

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. <https://doi.org/10.1093/poq/nfq058>

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, *23*(2), 532–544. <https://doi.org/10.1037/a0022402>

Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, *19*(4), 651–682. <https://doi.org/10.1177/1094428116656239>

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, *52*(4), 281–302.

Dienes, Z. (2008). *Understanding psychology as a science*. London: Palgrave Macmillan.

Donsbach, W. (2015). Public Opinion Polls. In G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication*. New York: Wiley-Blackwell.

Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In *Advances in experimental social psychology* (Vol. 33, pp. 41–113). <https://doi.org/10.1016/S0065-2601(01)80004-6>

Dumenci, L. (2000). Multitrait-Multimethod Analysis. In H. E. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 583–611). San Diego: Elsevier. <https://doi.org/10.1016/B978-0-12-691360-6.X5000-9>

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. <https://doi.org/10.1111/bjop.12046>

Fealy, G. (2014). Resurgent political Islam, or astute Islamic parties? *New Mandala*.

Fowler, F. J. (2014). *Survey research methods* (Fifth edition). Los Angeles: SAGE.

Fried, E. I., & Flake, J. K. (2018). Measurement Matters. *APS Observer*, *31*(3).

Groves, R. M., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, *74*(5), 849–879. <https://doi.org/10.1093/poq/nfq065>

Hunsley, J., & Meyer, G. J. (2003). The Incremental Validity of Psychological Testing and Assessment: Conceptual, Methodological, and Statistical Issues. *Psychological Assessment*, *15*(4), 446–455. <https://doi.org/10.1037/1040-3590.15.4.446>

Intan, G. (2019). Dituduh Tukang Bohong, Lembaga Survei Ungkap Data dan Metodologi Quick Count. *VOA Indonesia*. https://www.voaindonesia.com/a/dituduh-tukang-bohong-lembaga-survei-ungkap-data-dan-metodologi-quick-count/4885243.html.

Mair, P. (2018). *Modern Psychometrics with R*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-93177-7>

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*(4), 523–547.

McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods*, *23*(3), 412–433. <https://doi.org/10.1037/met0000144>

Messick, S. (1995). Validity of Psychological Assessment. *American Psychologist*, 9.

Mietzner, M. (2009). Political opinion polling in post-authoritarian Indonesia: Catalyst or obstacle to democratic consolidation? *Bijdragen Tot de Taal-, Land- En Volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, *165*(1), 95–126. <https://doi.org/10.1163/22134379-90003644>

Morling, B. (2018). *Research Methods in Psychology: Evaluating a World of Information* (3rd ed.). New York, NY: W. W. Norton & Company, Inc.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>

Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, *6*(MAR), 1–11. <https://doi.org/10.3389/fpsyg.2015.00223>

Peters, G. J. Y. (2018). *The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach’s alpha and the route towards more comprehensive assessment of scale quality* (Preprint). PsyArXiv. <https://doi.org/10.31234/osf.io/h47fv>

Revelle, W., & Zinbarg, R. E. (2008). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145. <https://doi.org/10.1007/s11336-008-9102-z>

Ruel, E. E., Wagner, W. E., & Gillespie, B. J. (2016). *The practice of survey research: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Sijtsma, K. (2008). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha. *Psychometrika*, *74*(1), 107. <https://doi.org/10.1007/s11336-008-9101-0>

Skrondal, A., & Rabe-Hesketh, S. (2007). Latent Variable Modelling: A Survey. *Scandinavian Journal of Statistics*, *34*(4), 712–745. <https://doi.org/10.1111/j.1467-9469.2007.00573.x>

Soderborg, S. (2019). Indonesia: How the polls are performing. *New Mandala*.

Thornton, S. (2019). Karl Popper. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two Quantitative Approaches for Estimating Content Validity. *Western Journal of Nursing Research*, *25*(5), 508–518. <https://doi.org/10.1177/0193945903252998>