Perbandingan Akurasi dan Efisiensi antara Survei dengan Kerangka *Complete listing dan Mapping*

Studi Kasus : Rumah Tangga Pengguna Internet di Kota Yogyakarta 2020

(*Comparison Accuracy and Efficiency between Survey with Complete listing Frame and Mapping Frame (Study Case : Internet Household in Yogyakarta Regency 2020)*)

M Irfan Zulfahmi^{1*}, Achmad Prasetyo²

^{1,2}Politeknik Statistika STIS

Jl. Otto Iskandardinata No. 64C, RT.01, RW 04, Bidara Cina, Kecamatan Jatinegara, Jakarta Timur, DKI Jakarta

E-mail: 211709799.stis.ac.id

ABSTRAK

Kegiatan pengumpulan data melalui survei skala besar sering dihadapkan pada kendala waktu, tenaga, dan biaya operasional, terutama apabila data yang akan dikumpulkan bersifat segera. Salah satu cara yang digunakan untuk mengatasi hal tersebut adalah dengan menggunakan survei rapid. Namun, survei rapid menggunakan kerangka sampel mapping yang mana tidak memuat seluruh unit observasi sehingga menghasilkan coverage error yang berpengaruh pada nilai estimasinya. Oleh karena itu, peneliti bertujuan untuk mengetahui nilai coverage error dan bias yang didapat kemudian membandingkan akurasi dan efisiensi antara survei dengan kerangka complete listing dan kerangka mapping. Tahapan penelitian berupa pembangunan kerangka, simulasi, perhitungan indikator, perbandingan RMSE, dan analisis efisiensi. Hasil estimasi pada kerangka mapping underestimate dan memiliki bias serta memiliki akurasi yang lebih rendah dibandingkan survei kerangka complete listing. Survei dengan kerangka mapping lebih efisien saat fix RMSE dan tidak lebih efisien dari survei dengan kerangka complete listing saat fix time.

Kata kunci: *Survei Rapid, Simulasi, Perbandingan, RMSE, Efisien*

ABSTRACT

Data collection activities through large-scale surveys are often faced with constraints of time, effort, and operational costs, especially if the data to be collected is urgent. One way to overcome this is to use a rapid survey. However, the rapid survey uses a mapping sample frame which does not contain all observation units, resulting in a coverage error that affects the estimated value. Therefore, the researcher aims to determine the coverage error and bias values observed and then compare the accuracy and efficiency of the rapid survey and conventional survey. The research's stage are development sampling frame, simulation, indicator calculation, RMSE comparison, and efficiency analysis. The estimation's results in the mapping frame is underestimated and has lower accuracy than conventional survey. Rapid survey are more efficient when fix RMSE and no more efficient than conventional survey when fix time.

Keywords: *Rapid Survey, Simulation, Compared, RMSE, Efficient*

PENDAHULUAN

Survei adalah metode sistematis untuk mengumpulkan informasi dari sampel sebuah entitas dengan tujuan untuk membangun karakteristik deskriptif kuantitatif dari atribut populasi yang lebih besar di mana entitas menjadi anggotanya (Groves, 2004). Keunggulan dari pengumpulan data survei adalah pada biaya yang lebih murah dan juga lebih cepat dibandingkan dengan sensus yang harus mengumpulkan populasi (semua) entitas. Oleh karena itulah, survei banyak dilakukan oleh banyak negara untuk pengumpulan datanya.

Meskipun survei memiliki biaya yang lebih murah dibandingkan dengan sensus, survei tetap memiliki biaya yang cukup mahal untuk menjaga kualitas dari survei itu sendiri. Selain biaya, survei juga terkendala pada masalah tenaga dan waktu karena kedua hal tersebut memiliki banyak hambatan ketika saat pengumpulan data di lapangan berlangsung. Untuk mengatasi hal tersebut, kegiatan survei berkembang dari pengambilan sampel yang sederhana hingga menuju yang kompleks mulai dari teknik pengambilan sampelnya maupun moda pengumpulan data yang digunakan.

Survei rapid merupakan salah satu metode pengumpulan data pada survei yang memiliki waktu yang lebih cepat dan juga biaya yang lebih murah dibandingkan dengan survei konvensional umumnya (Davis dkk., 2009). Survei rapid ini diinisiasi dan dikembangkan oleh PBB dan diterapkan lewat program survei rapid contohnya EPI (*Expanded Program on Immunization*), LQAS (*Lot Quality Assurance Sampling*), dan program

survei lainnya. Survei rapid ini juga banyak diberlakukan di berbagai negara di negara berkembang sebagai contohnya negara-negara di Afrika.

Survei rapid itu sendiri merupakan salah satu moda pengumpulan data yang menghilangkan tahapan pembuatan kerangka sampel *complete listing* (pendaftaran lengkap seluruh unit dalam klaster) dalam praktiknya dan menggantinya dengan kerangka sampel *mapping*. Dengan dihilangkannya tahapan *complete listing* pada kegiatan survei, biaya yang dikeluarkan dapat diminimalkan dibandingkan bila harus melakukan *complete listing*. Hal ini juga hemat dari segi waktu karena kegiatan *complete listing* yang menyita banyak waktu dapat dihilangkan sehingga membuat petugas dapat lebih cepat dalam menyelesaikan survei. Dengan keunggulan dapat mengurangi biaya dan mempercepat pelaksanaan survei itulah, survei rapid menjadi salah satu pilihan PBB dalam mengumpulkan data survei (Macintyre, 1999). Sama halnya yang dikatakan Davis dkk. (2009), survei rapid menyederhanakan *sampling design* dan juga cara pengumpulan data sehingga dapat menekan biaya dalam pelatihan maupun survei di lapangan serta mempermudah petugas sehingga kegiatan survei dapat berjalan dengan lebih cepat. Namun, Badan Pusat Statistik (BPS) selaku National Statistics Office belum mempertimbangkan survei rapid ini dalam pelaksanaan pengumpulan data. Hal ini dikarenakan kerangka sampel pada survei rapid tidak selengkap *listing* sehingga hasil estimasi dari survei rapid berpotensi menghasilkan bias. Bias sendiri merupakan jenis kesalahan yang dikarenakan perbedaan nilai harapan dugaan (estimator) yang tidak sama dengan nilai sebenarnya (Asra dan Prasetyo, 2015).

Mahasiswa Polstat STIS melalui Praktik Kerja Lapangan (PKL) 59 (2020) telah melakukan pengumpulan data dan juga penelitian terkait tingkat presisi yang dihasilkan survei rapid yang menggunakan kerangka *mapping*. Namun, penelitian yang dilakukan belum mampu menangkap bagaimana sisi *coverage error* yang dihasilkan oleh tidak lengkapnya kerangka sampel pada survei rapid. *Coverage error* ini tentunya tidak dapat diabaikan karena dapat menambah komponen bias dari estimasi yang dihasilkan dan menurunkan mutu dari statistik yang dihasilkan karena termasuk pada bagian *nonsampling error*. Oleh karena itulah, perlunya untuk mengetahui bagaimana hasil *coverage error* yang didapatkan dari hasil survei rapid yang mana menggunakan kerangka *mapping*.

Tentunya akan ada nilai *error* yang dihasilkan dari kerangka sampel yang tidak memuat seluruh populasi target akan menurunkan kualitas dari data itu sendiri pada dimensi akurasi data. Hal ini dikarenakan estimasi yang didapat akan memiliki bias karena kerangka sampel yang digunakan pada survei rapid tidak selengkap populasi target seperti *complete listing* melainkan menggunakan kerangka *mapping*. Kerangka *mapping* dibentuk melalui pendaftaran daftar bangunan rumah tangga pada jalan yang dipilih petugas lapangan secara *spinning bottle/ random walk* (Davis dkk., 2009). Walaupun pemilihan jalan dilakukan secara acak, tentunya perlu dikaji seberapa besar bias *coverage error* yang dihasilkan.

Untuk melihat apakah bias tersebut dapat ditoleransi atau tidaknya, perlu untuk melihat kembali pada konsep akurasi suatu data. Hal ini karena akurasi mencerminkan bagaimana total *error* (kesalahan) yang dihasilkan yang didapat dari komponen bias kuadrat dan varians yang dihasilkan. Dengan kata lain, total *error* dapat menunjukkan bagaimana perbedaan antara nilai suatu estimator terhadap populasi yang sebenarnya. Karena itu, ukuran akurasi dapat digunakan untuk membandingkan bagaimana kualitas dari penggunaan kerangka sampel yang berbeda. Oleh karena itu, akurasi kerangka *mapping* perlu untuk dibandingkan dengan akurasi kerangka *complete listing* yang umumnya dilakukan supaya dapat melihat performa dari masing-masing kerangka sampel.

Selain dari komponen bias, nilai *error* juga dapat dikurangi dengan menambah jumlah sampel sehingga komponen *sampling error* akan berkurang dan secara tidak langsung akan mengurangi total *error* dari suatu estimasi (Singh dan Chaudhary, 1986). Dari serangkaian cara yang dapat diterapkan diatas, perlunya penelitian yang lebih lanjut guna memperbaiki estimasi dari survei rapid sehingga total *error* yang dihasilkan dapat diminimalisasi.

Walaupun memiliki kelemahan dari sisi akurasi yang dikarenakan bias, survei rapid juga memiliki keunggulan dari sisi penghematan sumber daya baik waktu, tenaga, dan biaya yang lebih kecil karena ketiadaan komponen *complete listing* dalam penghitungan waktu, tenaga, dan biaya survei. Keunggulan ini dapat dimanfaatkan dengan mengalokasikan waktu, tenaga, dan biaya yang dihemat ke pengurangan *sampling error* yaitu dengan menambah jumlah sampel. Hal ini karena dengan menambah jumlah sampel dapat menurunkan varians dan secara tidak langsung akan mengurangi total *error* itu sendiri. Namun, perlunya dikaji secara lanjut berapakah banyaknya sumber daya yang dapat dihemat dan jumlah sampel optimum dari survei rapid. Hal ini dilakukan untuk mengetahui dengan pasti, apakah dengan total *error* yang sama, sumber daya dari survei rapid tetap lebih hemat dan, apakah dengan sumber daya yang sama, total *error* pada survei rapid lebih kecil. Oleh karena itulah, perlunya untuk mengetahui bagaimana efisiensi sumber daya yang dihasilkan dari survei rapid dibandingkan dengan survei konvensional.

Dari hal di atas, dapat dirumuskan tiga tujuan penelitian sebagai berikut. Tujuan pertama adalah mengetahui *coverage error* dan bias yang dihasilkan *coverage error* dari penggunaan kerangka *mapping*. Tujuan kedua adalah membandingkan akurasi yang dihasilkan antara kerangka *complete listing* dengan *mapping*. Tujuan ketiga yaitu mengetahui jumlah sampel optimal pada masing-masing kerangka dan menganalisis efisiensi yang dihasilkan.

Menurut Macintyre (1999), survei rapid dipandang sebagai salah satu pilihan untuk mengurangi biaya dalam survei dan menyediakan data dengan cepat untuk tujuan perancangan dan evaluasi kebijakan. Survei rapid telah dikembangkan dari bermacam-macam disiplin ilmu dan kebutuhan. Ekonomi pembangunan, studi pembangunan pedesaan, antropologi, statistik, epidemiologi dan kesehatan masyarakat dan gizi adalah beberapa disiplin ilmu yang telah membantu mengembangkan berbagai jenis survei rapid, dan dengan demikian, menurut definisi, masih banyak kontroversi tentang apa sebenarnya yang dimaksud dengan survei rapid. Namun, sebagian besar otoritas menyertakan lima atribut yang menjadi ciri survei rapid yaitu: biaya rendah, umpan balik cepat hasil, kuesioner pendek, ukuran sampel lebih kecil dari yang diharapkan, misalnya, dalam survei sampel tradisional, dan, semakin meningkat, pengambilan dan analisis data terkomputerisasi. Perbedaan lain yang secara umum disepakati dalam berbagai jenis survei rapid adalah survei kualitatif/kuantitatif. Banyak penulis tentang hal ini setuju bahwa jenis metode yang dipilih seharusnya hanya ditentukan oleh jenis pertanyaan yang diajukan.

Salah satu isu utama dalam survei rapid adalah tidak adanya *complete listing* yang bertujuan untuk mendapatkan kerangka sampel yang memuat seluruh daftar unit observasi dan *up to date*. Sedangkan pada survei rapid sendiri, tidak dilakukannya *complete listing* sehingga kerangka sampel yang lengkap dan relevan tidak bisa didapatkan. Sebagai gantinya, survei rapid melakukan proses *mapping* yang hanya mendaftarkan sebagian rumah tangga (*listing* terbatas). *Mapping* sebagian rumah tangga yang dilakukan pada survei rapid juga didasarkan prinsip *spinning bottle/pen*.

Davis dkk. (2009) menyebutkan bahwa *mapping* dengan prinsip *spinning bottle/pen* (*random walk*) dilakukan pada saat klaster telah diidentifikasi. Pusat geografis dari klaster itu dijadikan titik awal dalam melakukan *mapping* kemudian melakukan pemilihan jalan secara acak sambil mendata bangunan rumah tangga sampai menemui batas dari klaster tersebut. Bangunan yang sudah di-*mapping* kemudian dipilih secara acak sesuai dengan desain sampling yang digunakan.

Tidak lengkapnya kerangka sampel dapat menyebabkan *coverage error* (Asra dan Prasetyo, 2015). Groves dkk. (2004) juga mengungkapkan bahwa *coverage error* merupakan kesalahan yang diakibatkan karena adanya perbedaan antara target populasi dengan kerangka sampel. Groves dkk. (2004) juga mengungkapkan bahwa hal ini dapat mengakibatkan terjadinya bias yang merupakan salah satu komponen dari *non sampling error*.

Pada kasus survei konvensional, cara untuk mengatasi *coverage error* yang terjadi adalah dilakukannya *listing* secara lengkap (*complete listing*). *Complete listing* ini merupakan suatu kegiatan pendaftaran semua unit dalam populasi yang telah ditentukan terlebih dahulu dengan ciri identifikasinya (Asra dan Prasetyo, 2015). *Complete listing* dilakukan supaya mendapatkan kerangka sampel yang menurut Singh dan Chaudhary (1986) adalah daftar dari seluruh unit-unit sampel yang menentukan struktur pengambilan sampel dari suatu survei.

Dari hal tersebut dapat terlihat bahwa perbedaan umum antara survei rapid dan survei konvensional adalah terletak pada kerangka sampel yang digunakan. Survei konvensional umumnya menggunakan kerangka sampel yang diperoleh melalui *listing* lengkap (*complete listing*) sedangkan pada survei rapid menggunakan kerangka *mapping* yang merupakan hasil *random walk*. Dari kajian teori diatas, dapat disajikan beberapa kelebihan dan kelemahan kerangka sampel yang didapatkan melalui *complete listing* dan *mapping* sebagaimana yang ditunjukkan pada tabel 1 berikut.

Tabel 1. Kelebihan dan Kekurangan antara kerangka sampel *Complete listing* dan *Mapping*

<i>Complete listing</i>		<i>Mapping</i>	
Kelebihan	Kekurangan	Kelebihan	Kekurangan
Kerangka sampel lengkap	Waktu, tenaga, dan biaya yang lebih besar	Biaya, tenaga, dan waktu yang lebih hemat	Kerangka sampel tidak lengkap
Setiap unit <i>sampling</i> diobservasi di lapangan secara langsung sehingga data lebih valid		Setiap unit <i>sampling</i> diobservasi di lapangan secara langsung sehingga data lebih valid	

Namun, ukuran kebaikan survei tidak hanya dilihat dari segi bias yang dihasilkan namun pula pada *varians error* yang dihasilkan. Ukuran kebaikan survei yang melihat apakah suatu estimator dapat merepresentasikan parameter sebenarnya adalah akurasi (Wand dan Wang, 1996). Menurut Caldwell (2007), akurasi merupakan kombinasi dari presisi dan *trueness*. Presisi menyatakan berapa banyak variasi di dalam estimasi (antara sampel) sedangkan *trueness* menyatakan seberapa dekat estimasi yang dihasilkan dengan karakteristik populasi yang ingin diestimasi (parameter). Akurasi dapat diukur dengan *mean square error* (MSE) dengan presisi yang diukur melalui *varians* dan *trueness* diukur melalui bias. Selain MSE, RMSE (*root mean square error*) atau akar dari MSE secara ekuivalen juga dapat digunakan untuk melihat total *error* (Levy dan Lemeshow, 2008). Oleh karena itu, ukuran MSE dan RMSE dapat digunakan untuk menilai kualitas dan bahan perbandingan antara satu survei dengan survei lainnya.

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta}) \dots\dots\dots (1)$$

dimana:

$MSE(\hat{\theta})$ = Mean Square Error

$Var(\hat{\theta})$ = Varians

$Bias^2(\hat{\theta})$ = Bias Kuadrat

Nilai MSE tersebut dipengaruhi oleh *sampling error* dan *non sampling error* (Beimer dan Leiberg, 2003). Nilai *sampling error* dapat mempengaruhi nilai bias apabila menggunakan *probability sampling* atau penggunaan rumus yang tidak sesuai. Selain itu, nilai *sampling error* biasanya lebih berpengaruh pada *varians* yang diakibatkan tidaknya diambil seluruh populasi atau *sampling design* yang digunakan. Penggunaan *sampling design* bertahap dan jumlah sampel yang terlalu sedikit dapat membuat nilai *varians* lebih tinggi dibandingkan dengan penggunaan *sampling design* satu tahap dan jumlah sampel yang lebih banyak (Singh dan Chaudhary, 1986). Nilai *nonsampling error* dapat menyebabkan *systematic error* yang dapat menyebabkan bias dan *variable error* yang menyebabkan *varians* (Beimer dan Leiberg, 2003). *Systematic error* dapat dijabarkan berdasar sumber kesalahan yaitu kesalahan spesifikasi, kerangka *sampling*, *nonresponse*, pengukuran, dan *data processing*. Sedangkan *variable error* dapat bersumber pada kesalahan pengukuran dan *data processing*.

Selain dari sisi akurasi, dalam menentukan keputusan, desainer survei juga dihadapkan dengan sumber daya yang terbatas. Oleh karena itu, seorang desainer survei dihadapkan dalam menentukan titik optimal baik jenis *sampling design* yang digunakan maupun jumlah sampel yang diambil. Hal ini sesuai dengan apa yang dikatakan Levy dan Lemeshow (2008) bahwa kriteria survei yang baik adalah akurasi, biaya, dan fisibel dilakukan. Hal ini juga disampaikan oleh Beimer dan Leiberg (2003) bahwa harus ada keseimbangan antara biaya yang tersedia, total *error* (akurasi), dan dimensi kualitas lainnya. Oleh karena itu, keputusan letak titik optimal akan ditentukan pula berdasarkan dengan kendala yang dihadapi baik biaya, waktu, dan tenaga.

Umumnya, efisiensi suatu survei dilihat dari segi biaya. Namun, waktu juga dapat digunakan sebagai kendala dalam menentukan titik optimal. Hal ini karena didasarkan bahwa waktu dapat untuk menggambarkan biaya sesuai dengan apa yang dikemukakan Groves (2004). Hal ini didasarkan bahwa setiap desainer survei harus memenuhi dimensi kualitas *timeliness*. Data yang memenuhi kualitas *timeliness* berarti dapat menyediakan data yang *up to date* sehingga dapat memenuhi utilitas dari data memerlukan biaya yang besar (Wand and Wang, 1996). Oleh karena itu, terdapat korelasi antara biaya dan waktu sehingga waktu juga juga dapat digunakan untuk menggantikan biaya sebagai kendala survei.

Dalam menentukan suatu pilihan, kita dapat menetapkan nilai akurasi atau kendala. Melalui *fix accuracy* (Total Error Tetap), kita dapat memilih desain yang meminimumkan biaya, waktu, atau tenaga. Demikian sebaliknya, dengan *fix constraint* (kendala tetap), kita dapat memilih *design* yang meminimumkan MSE. Tentunya nilai minimum merupakan titik optimal yang sebaiknya dipilih desainer survei.

Berdasarkan kajian teori tersebut, dapat dibuat kerangka penelitian operasional sebagai berikut. Pengerjaan penelitian diawali dengan membentuk kerangka sampel *complete listing* dan *mapping*. Dari kedua kerangka tersebut, didapatkan nilai bias karena *coverage error* khusus pada kerangka *mapping* dan MSE dari kedua kerangka tersebut melalui simulasi *sampling*. Nilai RMSE yang dihasilkan kemudian dibandingkan dengan menggunakan uji beda dua rata-rata. Selain itu, dilihat pula titik jumlah sampel optimum dari penggunaan kedua kerangka sampel dan perbandingan efisiensi yang didapat.

Gambar 1. Kerangka Penelitian Operasional

METODE

Penelitian ini didahului *preprocessing data* dengan melihat blok sensus yang *eligible* untuk disimulasikan *random walk* melalui letak koordinat. Tahapan selanjutnya adalah dengan membentuk kerangka sampel *mapping*. Kerangka sampel *mapping* ini dibentuk dengan melakukan simulasi *random walk* yaitu dengan mencari titik tengah suatu klaster kemudian mendaftarkan rumah tangga yang dilalui jalan yang mana jalan tersebut dipilih secara acak. Selanjutnya, simulasi sampling pada kedua kerangka dengan menggunakan prinsip iterasi berulang (*monte carlo*) dengan *two stage sampling unequal first stage unit sampling design* dengan pengambilan sampel pada tahap pertama dan kedua diambil secara acak sederhana atau SRS. Ilustrasi skema sampling design dapat dilihat pada gambar 2.

Gambar 2. Skema Two Stage Sampling Unequal First Stage Unit

Tahap selanjutnya adalah melihat indikator yang dihasilkan dari simulasi berupa estimasi proporsi, bias, *varians sampling*, dan MSE dengan menggunakan jumlah sampel blok sensus $(n) = \{5, 6, 7, 8, 9\}$ dan jumlah sampel rumah tangga $(m) = \{5, 10, 15, 20, 25, 30, 35\}$. Pada tahapan ini, dilakukan analisis deskriptif dalam bentuk *boxplot* dan tabel. Tahapan selanjutnya adalah dengan melihat perbandingan antara RMSE yang dihasilkan dari penggunaan kerangka hasil *complete listing* dan *mapping*. Perbandingan dilakukan dengan melakukan analisis inferensia dengan perbandingan dua rata-rata populasi berbeda. Sebelum melakukan perbandingan, peneliti melihat menguji normalitas dengan menggunakan *Shapiro-wilk test* dan kesamaan

varians dengan menggunakan *Levene test*. Terakhir, peneliti melihat keputusan dengan mencari titik optimal dari penggunaan kedua kerangka sampel dengan kondisi *fix* RMSE dan *fix time* kemudian melihat efisiensi yang dihasilkan.

Pada kondisi *fix* RMSE, peneliti menetapkan RMSE yang diambil senilai 0.075. Hal ini diambil karena nilai dari RRMSE atau RMSE secara relatif (dalam persen) yang akan dihasilkan dengan variabel rumah tangga internet adalah 0.1104 atau 11,04%. Jika melihat RSE maksimal yang ditetapkan BPS yaitu 25%, maka RRMSE sebesar 11% dinilai cukup aman apalagi dalam hal ini RSE hanya memerhatikan nilai *standard error* sedangkan nilai RRMSE memerhatikan nilai *standard error* dan bias.

Pada kondisi *fix Time*, peneliti menggunakan standar waktu di PKL Polstat STIS T.A 2019/2020 yang memiliki konstrain waktu pencacahan selama satu minggu. Pada satu harinya, seorang pencacah memiliki waktu selama 8 jam. Namun, dalam melakukan pencacahan tidak semua 8 jam dapat dimaksimalkan. Dengan mengasumsikan bahwa 1 jam digunakan untuk istirahat makan siang dan 1 jam untuk mobilitas antara blok sensus dan antara rumah, sehingga pencacah akan memiliki waktu selama 6 jam khusus untuk melakukan pembentukan kerangka sampel dan wawancara. Dengan jumlah pencacah sebanyak 3 orang, maka waktu tetap yang digunakan adalah 7.560 menit (126 jam).

Data dan Sumber Data

Penelitian ini menggunakan data *complete listing* yang dilakukan di blok sensus terpilih di Kota Yogyakarta tahun 2020 yang dilakukan oleh mahasiswa politeknik statistika STIS. Oleh karena itu, populasi target pada penelitian ini hanya pada blok sensus terpilih. Variabel yang digunakan adalah rumah tangga internet yang merupakan variabel yang penting untuk melihat bagaimana distribusi internet suatu wilayah dan juga internet sebagai infrastruktur digital ekonomi. Untuk melihat efisiensi, peneliti menggunakan *constraint* berupa waktu total lapangan. Pada penelitian ini mengasumsikan nilai bias yang dihasilkan pada kesalahan spesifikasi, pengukuran, *nonresponse*, dan *data processing* serta nilai varians akibat variabel *error* karena kesalahan pengukuran dan *data processing* bernilai 0. Oleh karena itu, nilai MSE yang dihasilkan hanya memperhitungkan nilai bias akibat kesalahan kerangka sampel dan *varians sampling*.

HASIL DAN PEMBAHASAN

Sebelum dilakukan simulasi *sampling* dengan menggunakan prinsip *monte carlo* yang mana melakukan pengambilan sampel dan perhitungan indikator sebanyak n iterasi yang kemudian n indikator yang dihasilkan dirata-ratakan untuk mengetahui nilai ekspektasi dari indikator. Oleh karena itu, perlu untuk menentukan jumlah iterasi yang dilakukan. Jumlah iterasi yang ditentukan untuk menghasilkan indikator yang konvergen. Oleh karena itu, dilakukan simulasi untuk mengetahui jumlah iterasi optimal yang ditunjukkan pada gambar 3. Berdasarkan gambar tersebut, dapat dilihat bahwa semakin tinggi jumlah iterasi, maka nilai proporsi yang dihasilkan akan semakin konvergen dan semakin mendekati nilai proporsi sebenarnya (P). Nilai ekspektasi proporsi ($E[\hat{P}]$) konvergen pada jumlah iterasi sebanyak 25.000. Hal ini menandakan bahwa jumlah iterasi minimal yang dibutuhkan adalah sebanyak 25.000. Namun, peneliti memutuskan menggunakan 1.000.000 iterasi supaya hasil dari penelitian lebih dapat dipertanggungjawabkan.

Gambar 3. Jumlah Iterasi Optimal

Estimasi Proporsi

Nilai estimasi proporsi rumah tangga internet yang dihasilkan dengan menggunakan kerangka *complete listing* menghasilkan nilai berdasarkan *boxplot* yang ditunjukkan pada gambar 4. Dapat dilihat bahwa semakin

besar ukuran sampel maka rentang nilai proporsi yang dihasilkan pula akan semakin kecil. Rentang nilai proporsi yang dihasilkan itu pula mengindikasikan bahwa variabilitas yang dihasilkan akan semakin kecil dengan bertambahnya ukuran sampel. Namun, ukuran pemusatan data berupa median dari proporsi menunjukkan hasil yang mirip antara penggunaan ukuran sampel yang satu dengan yang lain.

Gambar 4. Estimasi Proporsi dengan *Complete listing*

Sedangkan untuk nilai estimasi proporsi rumah tangga internet yang dihasilkan dengan menggunakan kerangka *mapping* dapat dilihat melalui *boxplot* yang ditunjukkan pada gambar 5. Senada dengan kerangka *complete listing*, rentang dari hasil estimasi semakin kecil apabila ukuran sampel baik jumlah blok sensus ataupun rumah tangga yang diambil diperbesar. Hal ini menunjukkan bahwa variabilitas yang dihasilkan akan semakin kecil. Berbeda dengan rentang, nilai pemusatan data pada *boxplot* yaitu median dari sejumlah estimasi proporsi tidak mengalami perubahan yang signifikan dan cenderung stagnan apabila ukuran sampel diperbesar. Hal ini mengindikasikan bahwa ukuran sampel tidak memengaruhi nilai dari median sejumlah proporsi sampel.

Gambar 5. Estimasi Proporsi dengan *Mapping*

Terdapat perbedaan hasil ekspektasi dari estimasi proporsi yang dihasilkan dari penggunaan kerangka *complete listing* dengan kerangka *mapping*. Perbedaan tersebut dilihat gambar 6 berikut yang menunjukkan bahwa tidak ada gejala perpotongan antara nilai yang dihasilkan pada kedua kerangka. Nilai dari penggunaan *complete listing* memiliki nilai yang lebih besar dibandingkan dengan nilai yang dihasilkan dari penggunaan *mapping*. Dari grafik tersebut pula dapat dilihat bagaimana nilai estimasi yang dihasilkan dari kedua estimator dengan proporsi rumah tangga internet populasi. Nilai estimasi yang dihasilkan kerangka *complete listing* cenderung berdekatan dengan nilai populasi yang ditunjukkan dari garis estimasi kerangka *complete listing* yang berpotongan dengan nilai proporsi populasi. Sementara itu, nilai estimasi proporsi yang dihasilkan kerangka *mapping* berada dibawah nilai proporsi populasi yang menunjukkan bahwa estimasi proporsi rumah tangga internet yang menggunakan kerangka *mapping* merupakan *underestimate*.

Gambar 6. Perbandingan Estimasi Proporsi dengan *Complete listing* dan *Mapping* dengan Proporsi Populasi

Perbandingan Tingkat Akurasi

Dengan menggunakan jumlah sampel yang sama, dibandingkan tingkat akurasi yang dilihat dari total *error* yang didapatkan dari pengambilan sampel dengan menggunakan kerangka *complete listing* dengan kerangka *mapping*. Ukuran total *error* yang digunakan adalah RMSE (*Relatif Mean Square Error*).

Pada variabel rumah tangga internet, dapat dilihat perbedaan antara RMSE yang dihasilkan dari gambar 9. Dari hasil tersebut dapat dilihat bahwa nilai RMSE yang dihasilkan dari penggunaan kerangka *mapping* lebih tinggi dibandingkan dengan hasil estimasi dengan penggunaan kerangka *complete listing*. Oleh karena itu pada variabel rumah tangga internet yang dilihat secara nonformal, survei konvensional yang menggunakan kerangka *listing* memiliki akurasi yang lebih baik dibandingkan survei rapid yang menggunakan kerangka *mapping*.

Gambar 7. Perbandingan RMSE dengan menggunakan *Complete listing* dan *Mapping*

Selain itu, hasil pengujian normalitas dan kesamaan varians pada RMSE ditunjukkan pada tabel 7 berikut. Dari hasil tersebut dapat dilihat bahwa *p-value* dengan penggunaan kerangka *complete listing* maupun *mapping* lebih dari tingkat signifikansi yang digunakan. Oleh karena itu kesimpulan dari pengujian tersebut adalah tolak H_0 di mana RMSE rumah tangga internet baik yang menggunakan kerangka *complete listing* maupun *mapping* berdistribusi normal dengan tingkat signifikansi 5%. Selain itu, *p-value* pada *levane test* yang dihasilkan dari RMSE kerangka *complete listing* dan *mapping* rumah tangga internet lebih dari alfa. Hal ini memberikan kesimpulan gagal tolak H_0 yang menyatakan bahwa varians RMSE *complete listing* dengan varians RMSE *mapping* homogen pada tingkat signifikansi 5%.

Tabel 2. Uji Normalitas dan Kesamaan Varians

<i>Shapiro-wilk test</i>				
No.	RMSE	W-value	p-value	Keputusan
1	CL	0,9383	0,0819	Normal
2	M	0,9372	0,0766	Normal
<i>Levene Test</i>				
No.	RMSE	F-value	p-value	Keputusan
1	CL dan M	0,0689	0,7939	Varians Homogen

Dengan menggunakan hipotesis nol yakni tidak ada perbedaan RMSE yang dihasilkan dari penggunaan kerangka *complete listing* dan *mapping*, hasil pengujian dapat dilihat pada tabel 8. Pada variabel rumah tangga internet, *p-value* yang dihasilkan kurang dari alfa sehingga keputusan yang diambil adalah tolak H_0 . Oleh karena itu dengan tingkat signifikansi sebesar 5%, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara RMSE *complete listing* dibandingkan dengan RMSE yang dihasilkan dari penggunaan kerangka *mapping*.

Tabel 3. Uji Beda Dua Rata-rata

Variabel	t hitung	p-value 2 tailed	p-value 1 tailed
Ruta Internet	-2,211	0,03102	0,01551

Hasil tersebut sesuai dengan teori karena RMSE pada kerangka *mapping* memiliki bias yang jauh lebih besar dibandingkan dengan kerangka *complete listing*. Oleh karena itu, secara tidak langsung nilai kerangka *mapping* dapat memiliki RMSE yang lebih tinggi dibandingkan dengan kerangka *complete listing*. Hal ini ditunjukkan dengan nilai t hitung yang negatif dengan *p-value 1 tailed* sebesar 0,01551 (Tolak H_0) yang menunjukkan bahwa rata-rata RMSE kerangka *complete listing* lebih rendah dibandingkan dengan RMSE *mapping* dengan tingkat signifikansi sebesar 5%.

Fix RMSE, Minimum Time

RMSE menyatakan akar dari total *error* yang mungkin didapat dari suatu estimasi. Dalam penelitian ini, peneliti menggunakan RMSE kurang dari 0,075. Dengan nilai RMSE tersebut, peneliti mencari kombinasi jumlah sampel (n dan m) yang menghasilkan waktu yang paling sedikit. Dapat dilihat pada gambar 10, pada variabel rumah tangga internet, titik optimum untuk estimasi yang menggunakan kerangka *complete listing* menghasilkan titik optimum yang sama. Titik optimum saat menggunakan kerangka *complete listing* berada saat $n = 6$ dan $m = 10$ dengan waktu lapangan 5502 menit. Selain itu, titik optimum untuk estimasi variabel rumah tangga internet berada saat jumlah sampel $n = 7$ dan $m = 10$ dengan waktu lapangan 4823 menit. Pada titik optimum tersebut pula, waktu yang dapat dihemat dalam survei adalah sebesar 679 menit atau 11 jam 19 menit.

Gambar 8. Plot Keputusan

Fix Time, Minimum RMSE

Keputusan untuk memilih RMSE terendah dengan konstrain waktu tersebut dapat dilihat melalui gambar 8. Plot merah yang menyatakan titik penggunaan *complete listing* lebih rendah daripada titik biru *mapping*. Pada kondisi optimal, pilihan apabila menggunakan kerangka *complete listing* adalah kombinasi jumlah sampel $n = 8$ dan $m = 10$ dengan RMSE sebesar 0,06168. Sementara bila menggunakan kerangka *mapping*, pilihan yang diambil adalah kombinasi $n = 8$ dan $m = 15$ dengan RMSE sebesar 0,06277. Dengan waktu lapangan kurang dari 7560 menit, RMSE optimum pada estimasi dengan kerangka *complete listing* lebih rendah dibandingkan dengan RMSE optimum yang didapat dengan kerangka *mapping*.

Efisiensi

Hasil keputusan memang dapat diaplikasikan dalam menentukan keputusan. Namun tidak serta merta dapat menentukan kerangka mana yang lebih efisien karena nilai *fix* RMSE maupun *fix time* tidak tepat pada 0,075 maupun 7.650 menit. Oleh karena itu, perlu untuk menganalisis efisiensi di titik optimum dengan melihat perbandingan RMSE dikali dengan waktu *complete listing* dengan RMSE dikali waktu *mapping*. Hasil tersebut ditunjukkan pada tabel 3 berikut.

Tabel 3. Efisiensi ketika *Fix* RMSE ≤ 0.075 dan *Fix Time* ≤ 7560 Menit

		RMSE	Waktu	RMSE x Waktu	Efisiensi	Keputusan
Fix RMSE ≤ 0.075	CL	0,07155	5502	393,6681	1,1060	<i>Mapping</i> Lebih Efisien
	M	0,0738	4823	355,9374		
Fix Time $\leq$ 7560 menit	CL	0,06168	7336	452,48448	0,9805	<i>Complete listing</i> lebih Efisien
	M	0,06277	7352	461,48504		

Apabila dicermati dengan seksama plot pada gambar 8, dapat dilihat bahwa pada waktu lapangan yang relatif lebih pendek, kerangka *mapping* menghasilkan RMSE yang lebih kecil dibandingkan kerangka *complete listing*. Sementara pada waktu lapangan yang relatif lebih panjang, kerangka *complete listing* menghasilkan RMSE yang lebih kecil dibandingkan dengan kerangka *mapping*.

KESIMPULAN

Dari hasil dan pembahasan di atas, dapat ditarik kesimpulan dimana estimasi proporsi rumah tangga internet yang dihasilkan dengan menggunakan kerangka *mapping* memiliki nilai yang *underestimate* sehingga menghasilkan estimator yang bias. Selain itu, akurasi yang dihasilkan dari penggunaan kerangka *complete listing* lebih baik dibandingkan dengan akurasi yang dihasilkan dengan kerangka *mapping*. Terakhir, pada kondisi *fix* RMSE, survei dengan menggunakan kerangka *mapping* lebih efisien dari sisi waktu dibandingkan dengan kerangka *complete listing* sedangkan, pada kondisi *fix time*, kerangka *complete listing* lebih efisien dibandingkan dengan kerangka *mapping*.

DAFTAR PUSTAKA

- Asra, Abuzar & Prasetyo, Achmad. (2015). *Pengambilan Sampel : Dalam Penelitian Survei*. Jakarta : RajaGrafindo Persada.
- Biemer, Paul P. & Lyberg, Lars E. (2003). *Introduction to Survey Quality*. New Jersey : John Wiley & Sons.
- Caldwell, Joseph G. (2007). *Sample Survey Design and Analysis : A Comprehensive Three-Day Course with Application to Monitoring and Evaluation*. Spartanburg : Foundation.
- Davis, R., Luna, J., Rodriguez-Lainz, A., & Sarriot, E. (2009). *The Rapid Household Survey: How to Obtain Reliable Data on Health at the Local Level*. California: Public Health Institute and ICF Macro.
- Groves, Robert M. (2004). *Survey Errors and Survey Costs*. New Jersey : John Wiley & Sons.
- Groves, Robert M., Fowler-Jr, Floyd J., Couper, Mick P., Lepkowski, James M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. New Jersey: John Wiley & Sons.
- Kish, Leslie. (1965). *Survey Sampling*. New York: John Wiley & Sons.
- Levy, Paul S. & Lemeshow, Stanley. (2008). *Sampling of Populations : Methods and Applications (4th ed)*. New Jersey: John Wiley & Sons.
- Macintyre, Kate. (1999). Rapid assessment and sampel surveys: trade-offs in precision and cost. *Health Policy and Planning*, 14(4), 363-373.
- PKL 59. (2020). *Perbandingan Metode Pengumpulan Data Survei Konvensional dan Survei Rapid pada Survei Terintegrasi SDGs Pilar Lingkungan dan Ekonomi Digital*. Jakarta : Politeknik Statistika STIS.
- Singh, Daroga & Chaudhary, Fauran S. (1986) *Theory and Analysis of Sample Survey Designs*. New York: John Wiley & Sons.
- Wand, Yair & Wang, Richard Y. (1996). Anchoring Data Quality Dimension in Ontological Foundations. *Communications of the ACM*, 39(11), 86-95.