

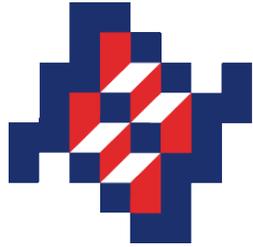


EL SISTEMA DE EVALUACIÓN DEL DESARROLLO SOCIAL DE LA CIUDAD DE MÉXICO:

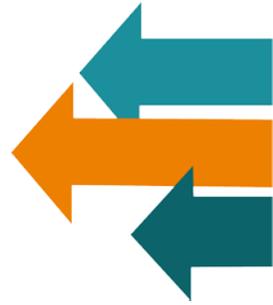
**HERRAMIENTAS PARA LA ELABORACIÓN
DE LAS EVALUACIONES INTERNAS 2018
DE LOS PROGRAMAS SOCIALES**

LEVANTAMIENTO DE INFORMACIÓN CUANTITATIVA





Instituto Nacional
de Salud Pública



Las encuestas como herramienta de generación de información cuantitativa

Juan Pablo Gutiérrez

 @gutierrezjp



Esquema de evaluación integral



Evaluación de resultados

Medición de los indicadores identificados para un programa:

- Insumos, Proceso, Producto, Resultados (intermedios o finales)

Medición del desempeño

- Eficiencia: lograr el resultado con el menor consumo de insumos
- Eficacia (efectividad): lograr el resultado
- Calidad: dar respuesta a los usuarios
- Economía: uso adecuado de recursos

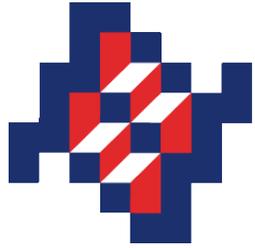


Métodos de obtención de información

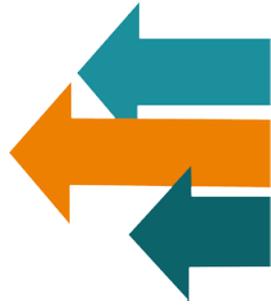
Registros administrativos

Formatos de obtención de información





Instituto Nacional
de Salud Pública



Obtención de información primaria: Encuestas

¿Para qué sirven las encuestas?

Las encuestas permiten obtener información diversa sobre un determinado grupo o fenómeno

- Condiciones de salud / bienestar
- Uso de tiempo
- Nivel socioeconómico
- Comportamientos
- Calidad de servicios

Genera datos de fuentes directas/primarias y de manera oportuna



¿Qué es una encuesta?

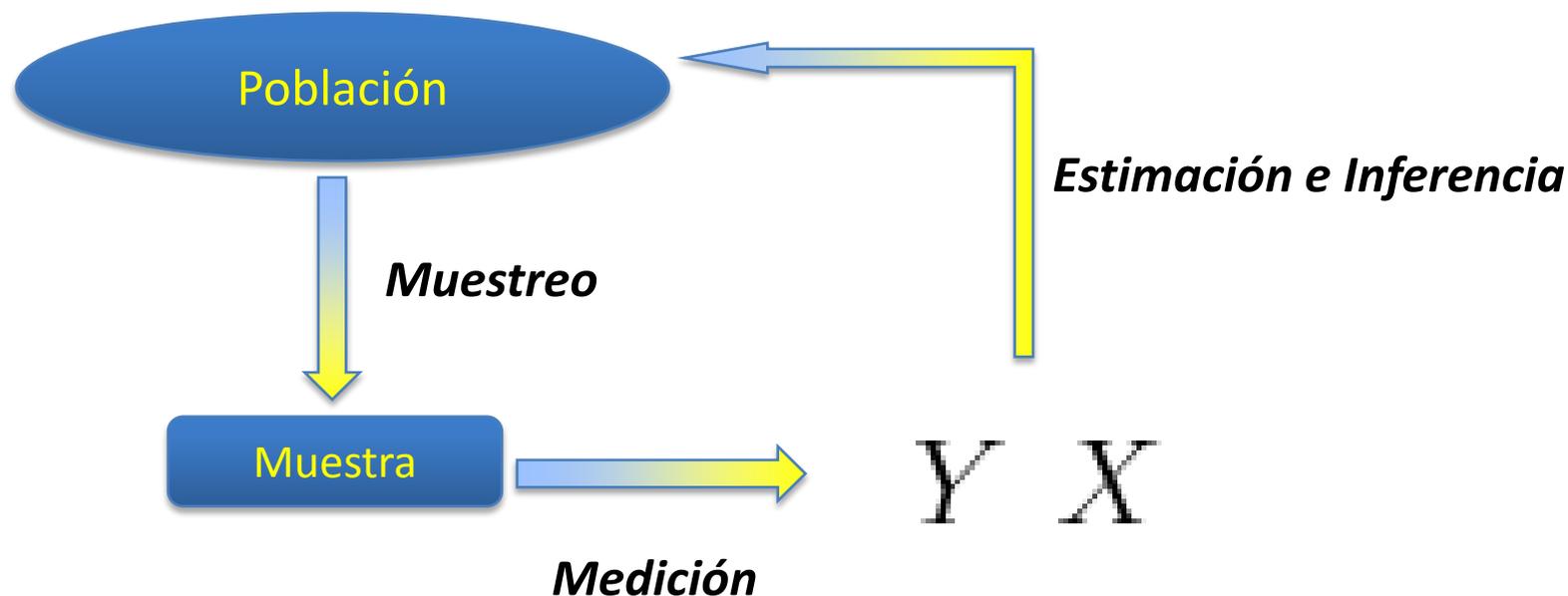
Técnica de investigación cuantitativa

Información de una muestra probabilística

Económicas, prácticas y rápidas



Elementos de una encuesta

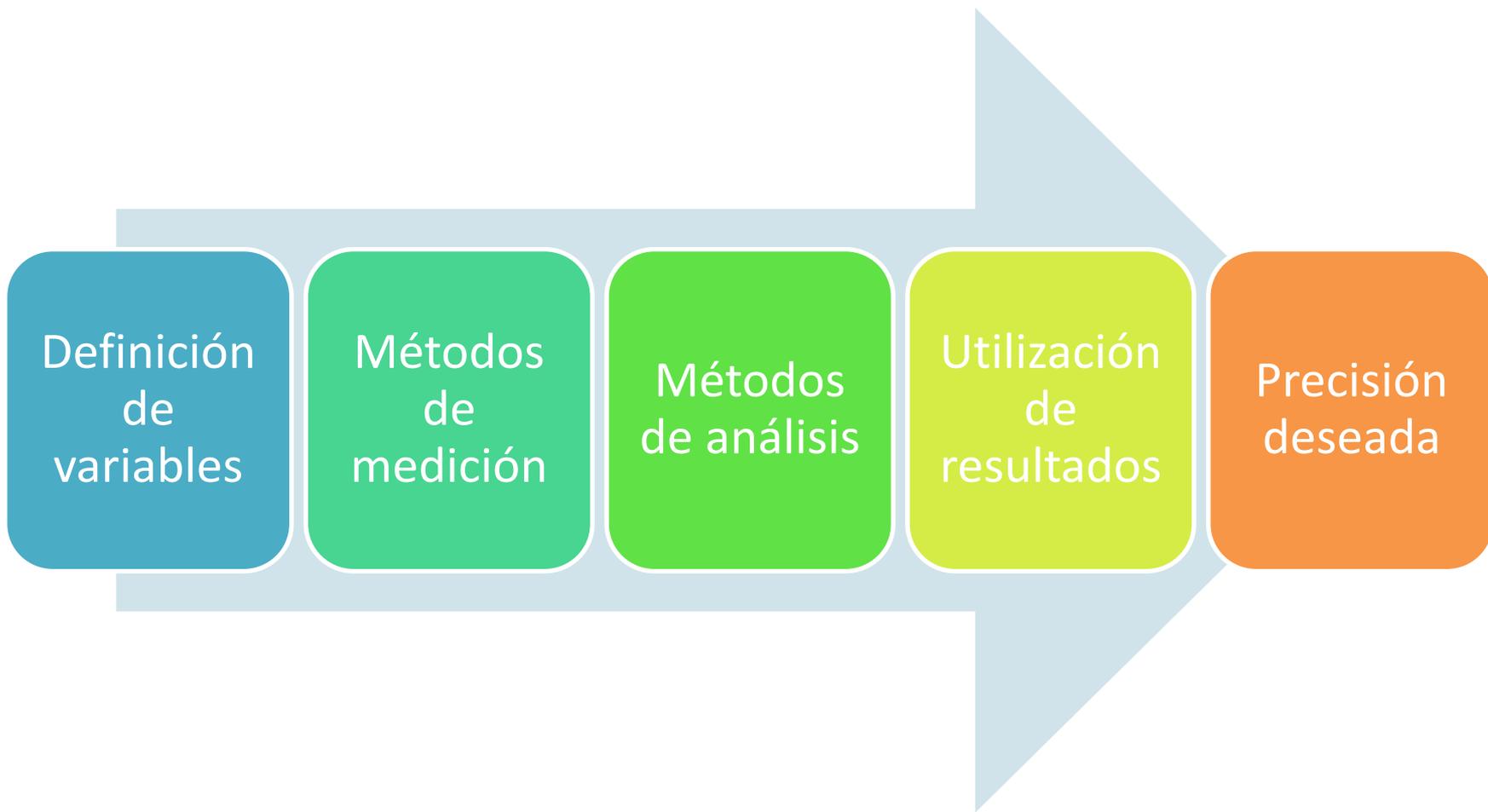


Población : Habitantes de viviendas particulares habitadas

Muestra : Elementos de la población seleccionados con probabilidades conocidas



Diseño de encuestas



Aspectos del diseño de muestras

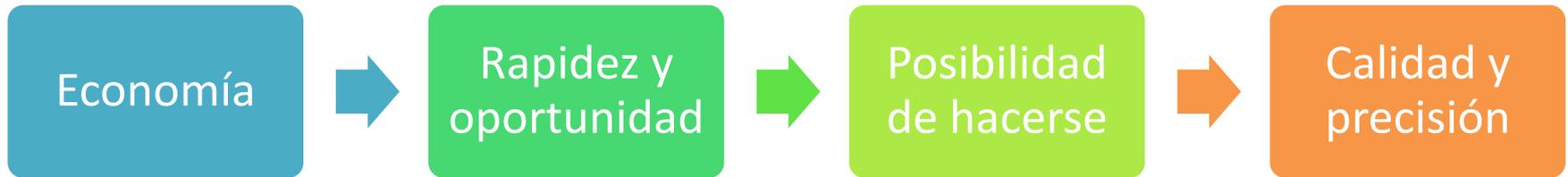
Proceso de selección

Proceso de estimación

El objetivo del diseño de muestras es la elección del diseño con el menor error



Ventajas de una muestra



Muestreos no probabilísticos

Muestras casuales

Selección experta

Muestreo de cuotas

Muestreo de poblaciones móviles



Taxonomía de métodos probabilísticos

Aleatorio simple

Sistemático

Estratificado

De una etapa o polietápico

Por conglomerados



Retomando, entonces, ¿por qué hacemos encuestas?

Imposibilidad de obtener información de toda la población en todo momento

- Muestra de individuos / hogares / establecimientos
- Información obtenida en un momento en el tiempo

¿Alternativas?

- Registros administrativos completos y confiables



¿Por qué muestras complejas?

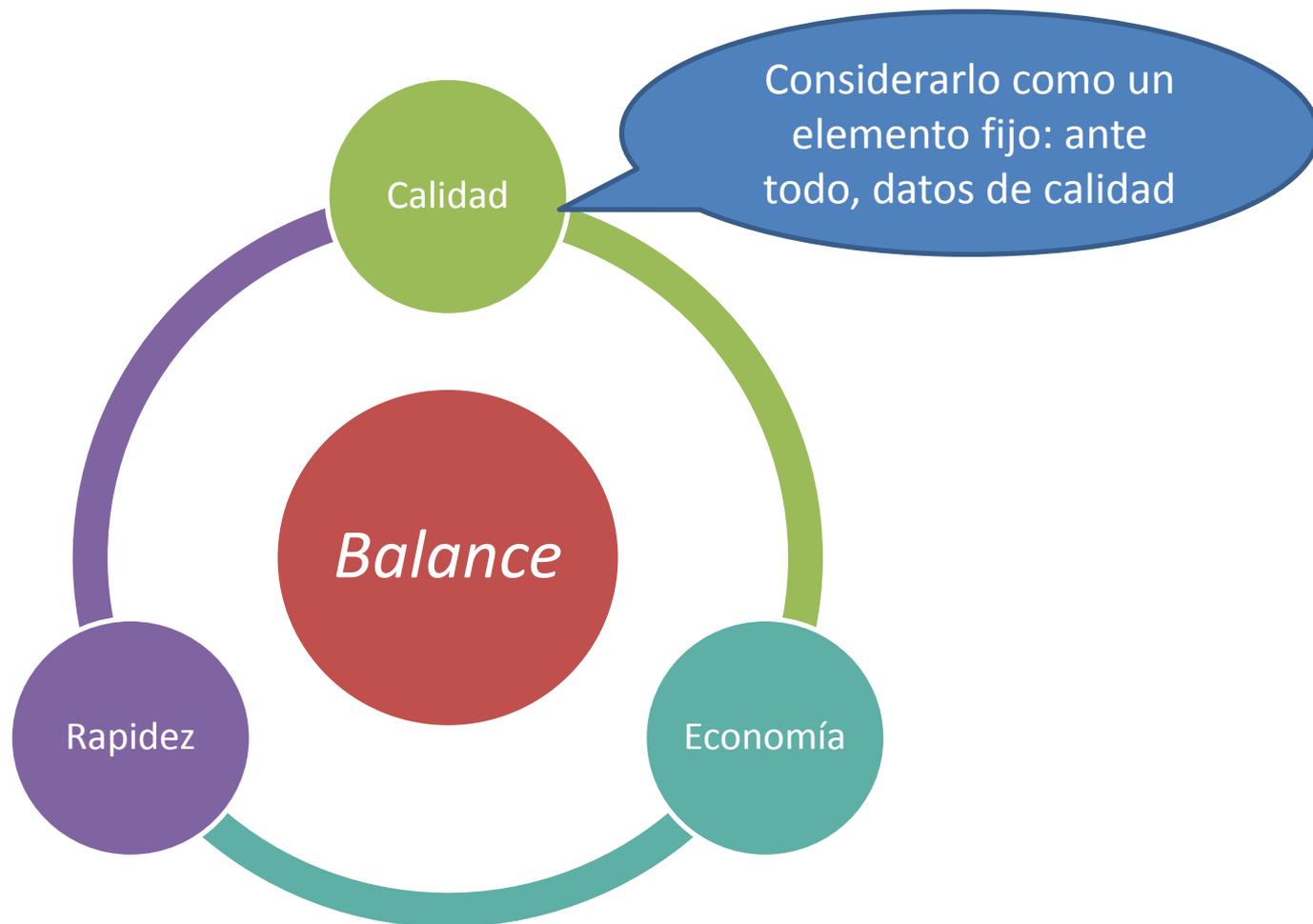
¿Factibilidad de selección aleatoria simple?

- Alta para una población acotada
- Baja para una población distribuida en un territorio amplio

¿Qué elementos inciden para una encuesta compleja?



Uno de los tres se ve afectado



Un paréntesis

PROCESO PARA PRESUPUESTAR UNA ENCUESTA



Instituto Nacional
de Salud Pública



Entender las necesidades del cliente permita generar estimaciones sobre los costos de recolección de información

Identificar las necesidades

- ¿Unidad de análisis?
- ¿Representatividad buscada?
- ¿Indicadores solicitados?

¿Qué productos se esperan?

- Diseño de la encuesta
- Base de datos
- Análisis de la información



Una estimación apropiada de los recursos necesarios permite certidumbre en el proyecto

Sobre-presupuestar disminuye las posibilidades de realizar el proyecto

Sub-presupuestar disminuye la calidad del proyecto

- Reducir tramo de control
- Menor tasa de respuesta
- Incumplimiento



En el punto más básico...

$$\begin{array}{|c|} \hline \text{Cantidad de} \\ \text{Insumos} \\ \hline \end{array} \times \begin{array}{|c|} \hline \text{Costos} \\ \text{Unitarios} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{Costos Totales} \\ \hline \end{array}$$



¿Cuáles son los parámetros para calcular el costo de una encuesta?

Factores que determinan los costos de forma directa

- Duración del instrumento: incrementa tiempo/entrevistador
- Tamaño de muestra: incrementa tiempo/entrevistador

Factores que pueden modificar los costos

- Tiempo factible en campo
 - En general, hasta un cierto límite, más tiempo en campo permite reducir costos
- Tramos de control
 - Mayor control incrementa los costos por personal de supervisión



Hay otros factores que afectan la estimación de los costos

Otros parámetros necesarios

- Horas efectivas de trabajo por día
- Días efectivos de trabajo por mes (o por semana)
- Tasa de respuesta
 - Modifica el tiempo promedio por observación (reduce costos)
 - Afecta el tiempo necesario de re-visitas (incrementa costos)
- Gastos de campo con pernocta o sin pernocta
 - Equipo de campo del lugar de trabajo o brigadas de trabajo
- Tiempo para croquis y listados (cartografía)



Considerar casos especiales

Encuestas con auto-aplicación

Establecimientos

- Cárceles

¿Otros?



Retos de muestras complejas

Balance entre precisión y costos

Los individuos en un conglomerado tienden a ser más parecidos entre ellos que con los de otros conglomerados:

- 10 hogares del mismo conglomerado nos dan menos información que 10 hogares de conglomerados diferentes

Consecuencia: errores estándar mayores

La pérdida de precisión puede ser anticipada en el diseño incrementando el tamaño de muestra

Efecto de diseño (deff)

- Necesario cuando se considera la precisión de muestras con cluster
- Mide el desempeño de un método de muestreo comparado con una muestra aleatoria simple (SRS)

$$deff = \frac{\text{Var}_{cluster}}{\text{Var}_{srs}}$$

- *deff* puede ser estimado tanto para medidas de asociación como para medidas de prevalencia (o medias)



Correlación intraconglomerado (ρ)

- d_{eff} depende del tamaño del conglomerado y de la fuerza de la correlación al interior de los

$$D_{eff} = [1 + \rho(k - 1)]$$

donde ρ = correlación intraconglomerado

k = tamaño del conglomerado

$$\rho = \frac{d_{eff} - 1}{k - 1}$$

d_{eff} = efectodiseño

k = tamaño de conglomerado

- ρ es una medida del grado de homogeneidad entre los sujetos del conglomerado para la medida de interés

Tamaño de muestra

- Usar métodos estándar de estimación de muestra bajo muestreo aleatorio simple, y después calcularlo usando un valor de *deff*

$$n = n_{srs} \times deff$$

where n_{srs} = size required with simpler random sampling

- Se necesita el tamaño estimado del conglomerado y un valor de ρ - que es una característica de la población estudiada

Retos con el *deff*

El valor de *deff* difiere para cada indicador y, en realidad, para cada grupo meta, ya que la homogeneidad de conglomerados varía por característica (Manual MICS)

Los valores de los *deffs* generalmente no se conocen para indicadores antes de la encuesta, pero se espera que sean bastante pequeños para varios indicadores

Usar información de fuentes similares

- (Manual MICS)



Máxima

ANALIZAR RESPETANDO EL DISEÑO



¿Qué ocurre si se hace de otra forma?

Supuestos
al no
considerar
el diseño

- Se trata de una muestra aleatoria simple, lo que implica que es una muestra auto-ponderada
- Esto es, cada unidad **en la base de datos** tuvo la misma probabilidad de selección
- Los estratos de interés en la muestra representan proporcionalmente a los estratos de interés en la población
- No hay efecto de diseño



¿Qué debe considerarse para el análisis?

Ponderación

- Ponderadores iniciales de selección (inverso de probabilidad de selección)
- Corrección de ponderadores por no respuesta

Diseño

- Estratos de selección
- Unidades primarias de muestreo (usualmente las encuestas poblacionales son por conglomerados)

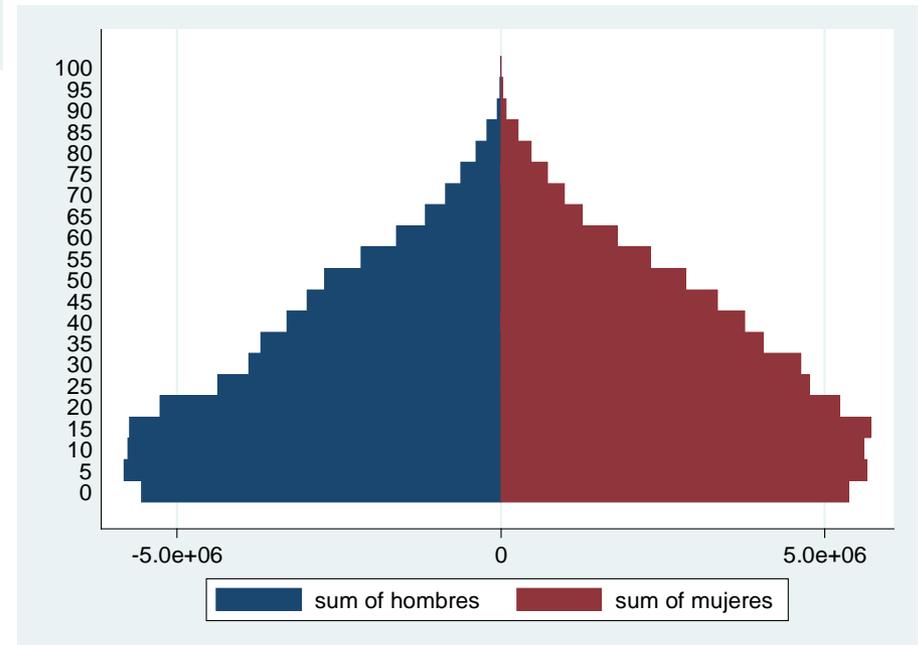
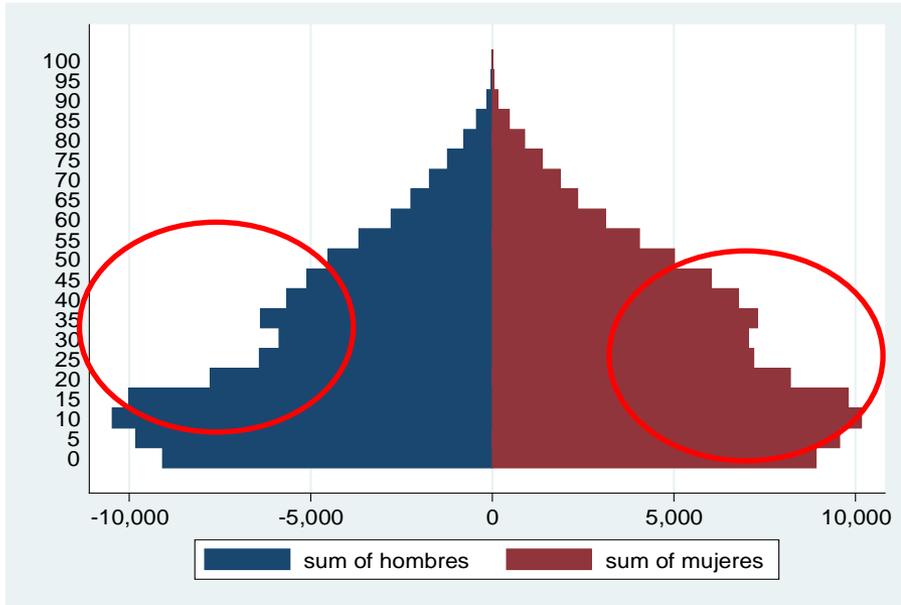


Importancia de los ponderadores

- Gráfica 1:
 - `graph hbar (sum) hombres mujeres [pw=pondei], over (gpo_edad, descending gap(0)) stack`
- Gráfica 2:
 - `graph hbar (sum) hombres mujeres, over (gpo_edad, descending gap(0)) stack`



Gráfica 1. Sin ponderadores



Gráfica 2. Con ponderadores



Ejemplo en Stata

- use "\$bases_org\integrantes.dta", clear;
- svy:prop afilia_1ra;
- tab afilia_1ra;



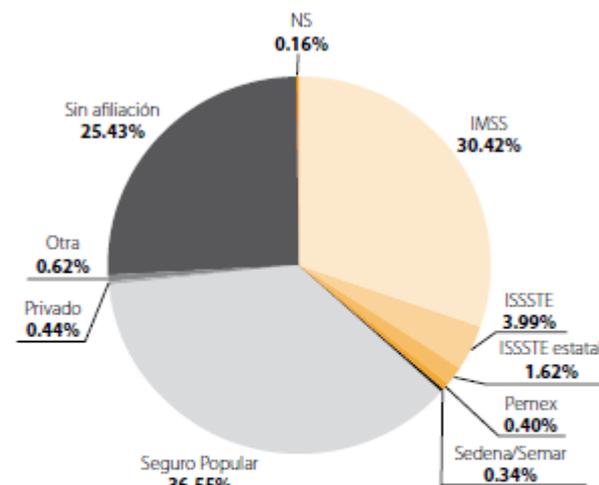
Estos son los porcentajes en la figura 1.2 del reporte de la ENSANUT 2012

Survey: Proportion estimation

Number of strata = 155 Number of obs = 194923
 Number of PSUs = 1622 Population size = 115170278
 Design df = 1467

_prop_5: afilia_lra = Defensa/Marina
 _prop_6: afilia_lra = Seguro Popular
 _prop_9: afilia_lra = NS/NR

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
afilia_lra				
Ninguna	.2543602	.0034141	.2476632	.2610571
IMSS	.3041183	.0046824	.2949334	.3133032
ISSSTE	.056177	.0018097	.0526271	.0597269
Pemex	.004065	.000691	.0027096	.0054205
_prop_5	.0034561	.0004774	.0025197	.0043925
_prop_6	.3655547	.0044877	.3567517	.3743577
Privado	.0044142	.0004453	.0035408	.0052877
Otro	.0062463	.0005346	.0051977	.007295
_prop_9	.0016082	.0001822	.0012507	.0019656

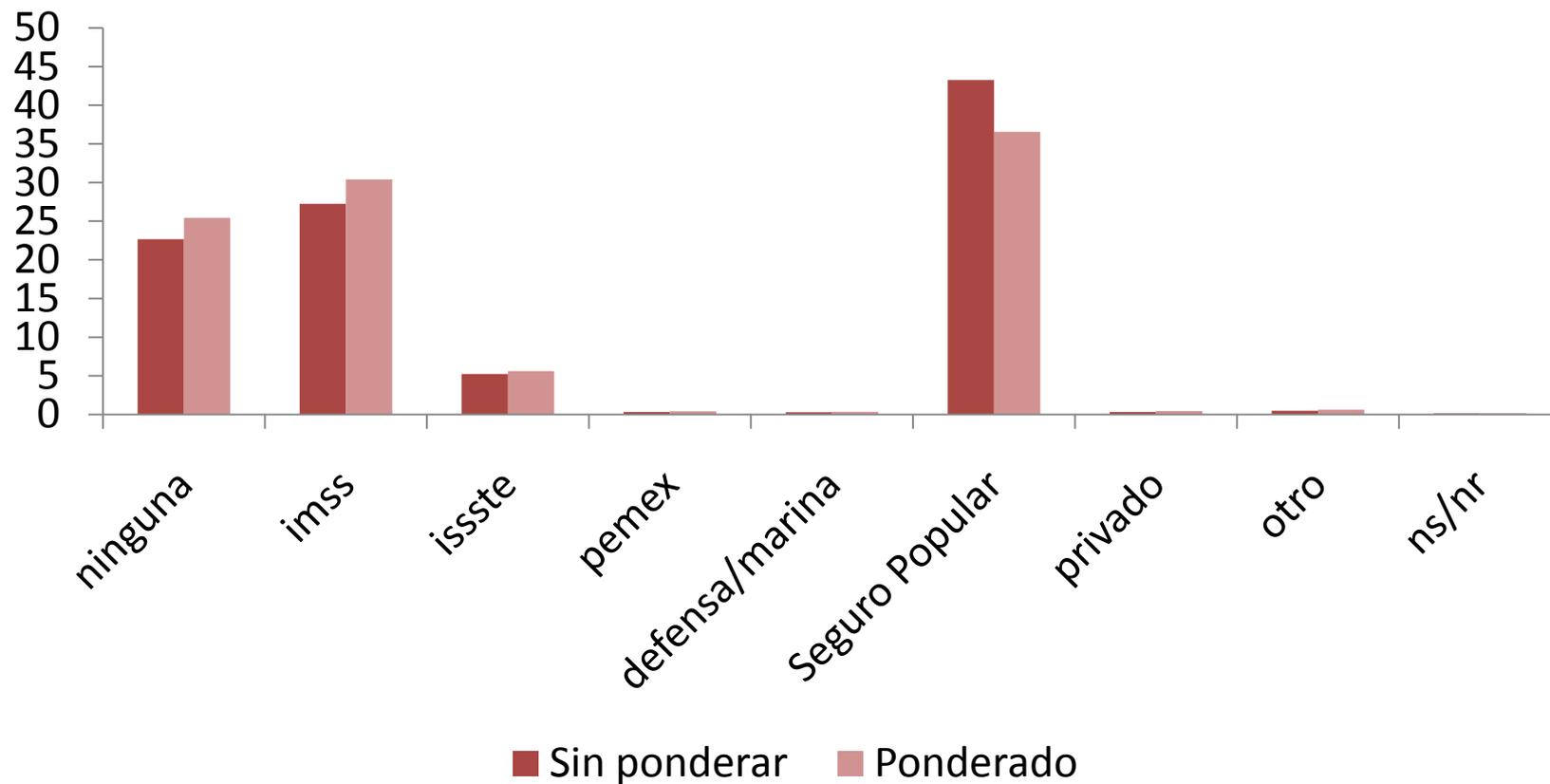


NS= No sabe

Figura 1.2
 Distribución de la población según institución de protección en salud de acuerdo con el autorreporte del informante del hogar. México, ENSANUT 2012



Diferencia por la ponderación



Diferencia por el diseño

```
. svy:prop afilia_lra;
(running proportion on estimation sample)
```

Survey: Proportion estimation

```
Number of strata =    155
Number of PSUs   =   1,622
Number of obs    =   194,923
Population size  = 115,170,278
Design df       =    1,467
```

```
_prop_5: afilia_lra = defensa/marina
_prop_6: afilia_lra = seguro popular
_prop_9: afilia_lra = ns/nr
```

```
. svy:prop afilia_lra
(running proportion on estimation sample)
```

Survey: Proportion estimation

```
Number of strata =    1
Number of PSUs   = 194,923
Number of obs    =   194,923
Population size  = 115,170,278
Design df       =   194,922
```

```
_prop_5: afilia_lra = defensa/marina
_prop_6: afilia_lra = seguro popular
_prop_9: afilia_lra = ns/nr
```

	Linearized			
	Proportion	Std. Err.	[95% Conf. Interval]	
afilia_lra				
ninguna	.2543602	.0034141	.2477215	.261115
imss	.3041183	.0046824	.2950122	.3133805
issste	.0561612	.0018096	.0527153	.0598181
pemex	.004065	.000691	.0029118	.0056724
_prop_5	.0034561	.0004774	.0026355	.004531
_prop_6	.3655705	.004488	.3568126	.3744182
privado	.0044142	.0004453	.0036215	.0053796
otro	.0062463	.0005346	.0052805	.0073875
_prop_9	.0016082	.0001822	.0012876	.0020084

	Linearized			
	Proportion	Std. Err.	[95% Conf. Interval]	
afilia_lra				
ninguna	.2543602	.0014243	.2515787	.2571618
imss	.3041183	.0014896	.3012065	.3070458
issste	.0561612	.0007443	.0547201	.057638
pemex	.004065	.0002044	.0036834	.004486
_prop_5	.0034561	.0002065	.003074	.0038854
_prop_6	.3655705	.0014585	.3627167	.3684338
privado	.0044142	.0002309	.0039841	.0048907
otro	.0062463	.0002698	.0057391	.006798
_prop_9	.0016082	.0001418	.001353	.0019114



Datos ENSANUT MC 2016

- Representatividad para el país y regional
- Análisis sobre calidad de atención a individuos adultos que reportaron algún padecimiento crónico



¿En dónde se atiende la población?

```
. tab lugar_atencion
```

lugar_atencion	Freq.	Percent	Cum.
Seg_Soc	918	37.39	37.39
SP/Prospera	1,068	43.50	80.90
Privado	469	19.10	100.00
Total	2,455	100.00	



¿En dónde se atiende la población?

```
. svy:prop lugar_atencion  
(running proportion on estimation sample)
```

Survey: Proportion estimation

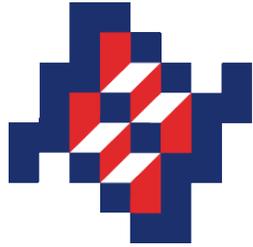
```
Number of strata =      39      Number of obs   =      2,455  
Number of PSUs   =      323     Population size = 18,021,200  
                                   Design df       =      284
```

```
_prop_2: lugar_atencion = SP/Prospera
```

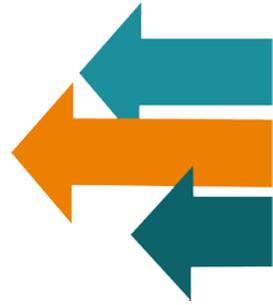
	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
lugar_atencion				
Seg_Soc	.4066524	.0228006	.3626633	.4521915
_prop_2	.3292647	.0184668	.2939864	.3665783
Privado	.2640829	.0227044	.2218687	.3111177

Note: Strata with single sampling unit centered at overall mean.





Instituto Nacional
de Salud Pública



Las encuestas como herramienta de generación de información cuantitativa

Juan Pablo Gutiérrez

 @gutierrezjp