



Animal Health Matters.  
For Safe Food Solutions.



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,  
Education and Research EAER  
**State Secretariat for Economic Affairs SECO**



# Ознайомлення з базовою концепцією відбору проб і розміру вибірки

Марко Де Харді



# Активне спостереження



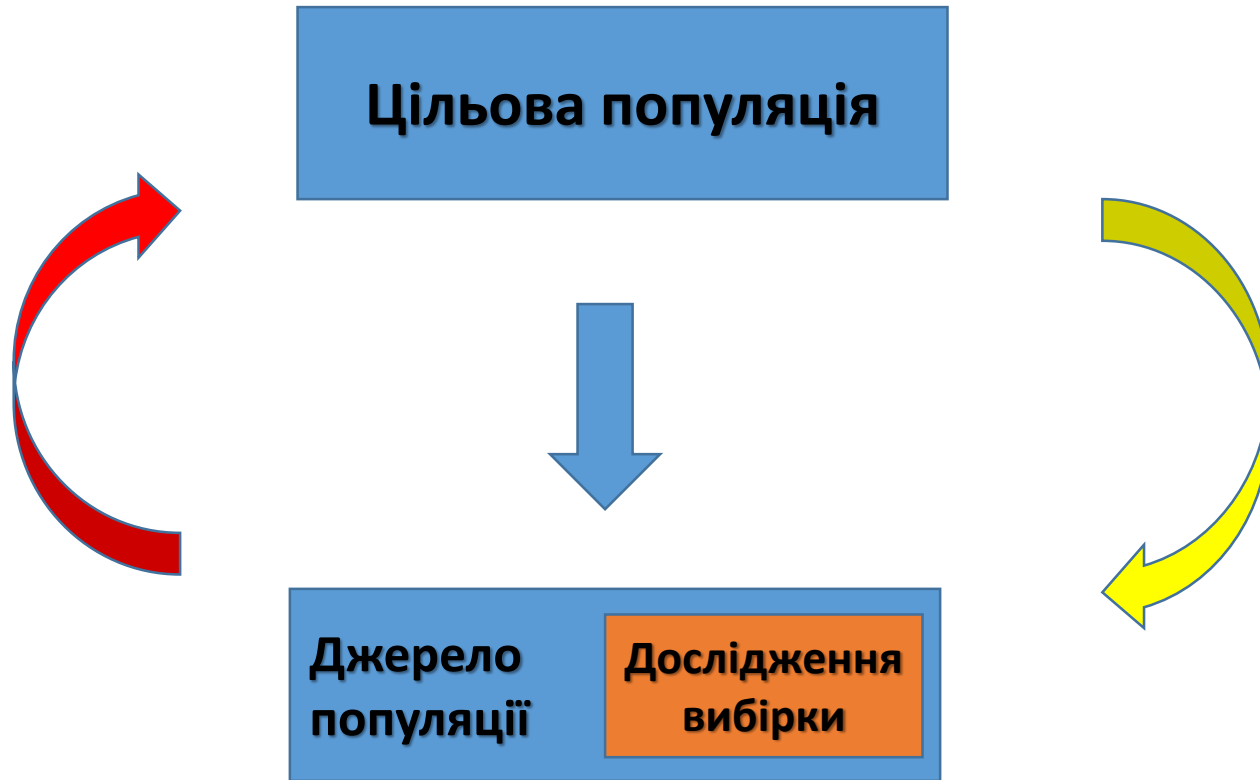
- **Різноманітність методів :**

1. Вичерпне спостереження / перепису
2. Спостереження за репрезентативною вибіркою
3. Ризик-орієнтоване/Цільове спостереження
4. Нагляд
5. Комбінація методів:



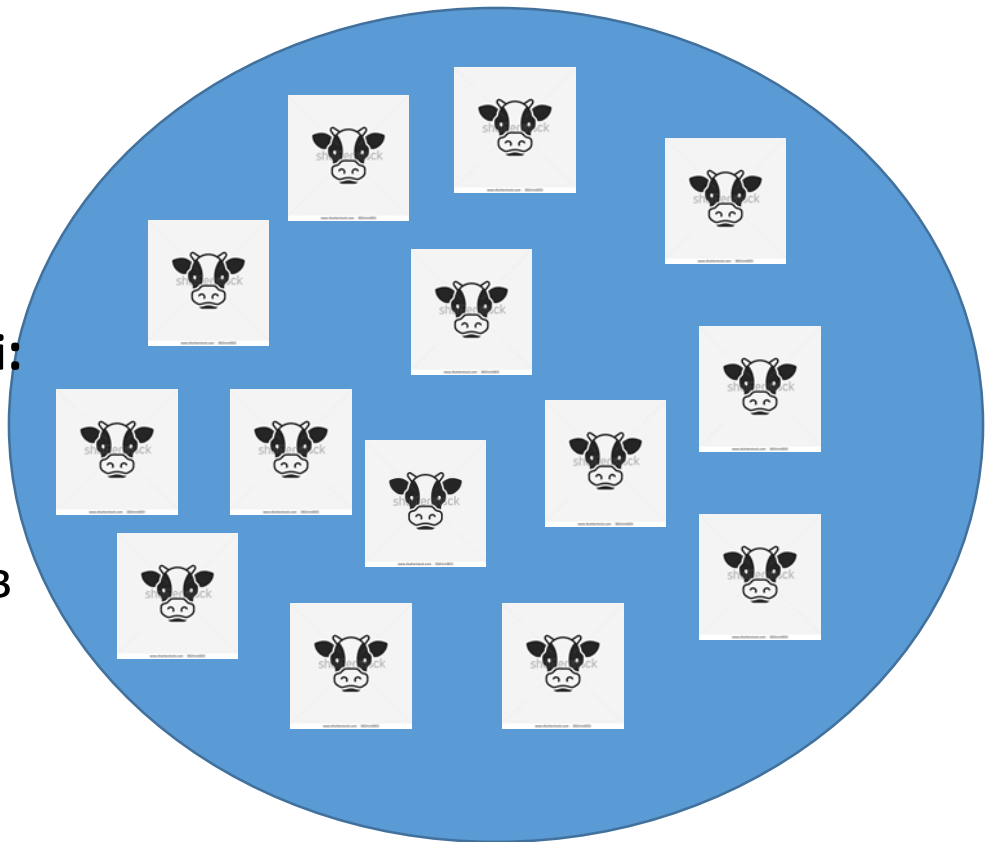


# Цільова популяція, джерело популяції, дослідження вибірки



# Цільова популяція

- Це популяція, на яку націлене спостереження, і про яку необхідна інформація
  - Не завжди чітко визначена (може бути дуже важко)
  - Відноситься до **цілої** групи осіб або предметів, по відношенню до яких дослідники зацікавлені узагальнити висновки
  - В ідеалі це має бути популяція в зоні ризику
- 
- **Мастити голштинської породи на фермах середніх розмірів в Україні:**  
Цільова популяція=  
усі корови голштинської породи на фермах середніх розмірів в країні



# Джерело популяції

- Це різновид (підгрупа) популяції
- Це популяція, з якої відбираються особини
- Всі індивіди (тварини, ферми) повинні "перевірятися" і мати ненульову ймовірність включення в дослідження

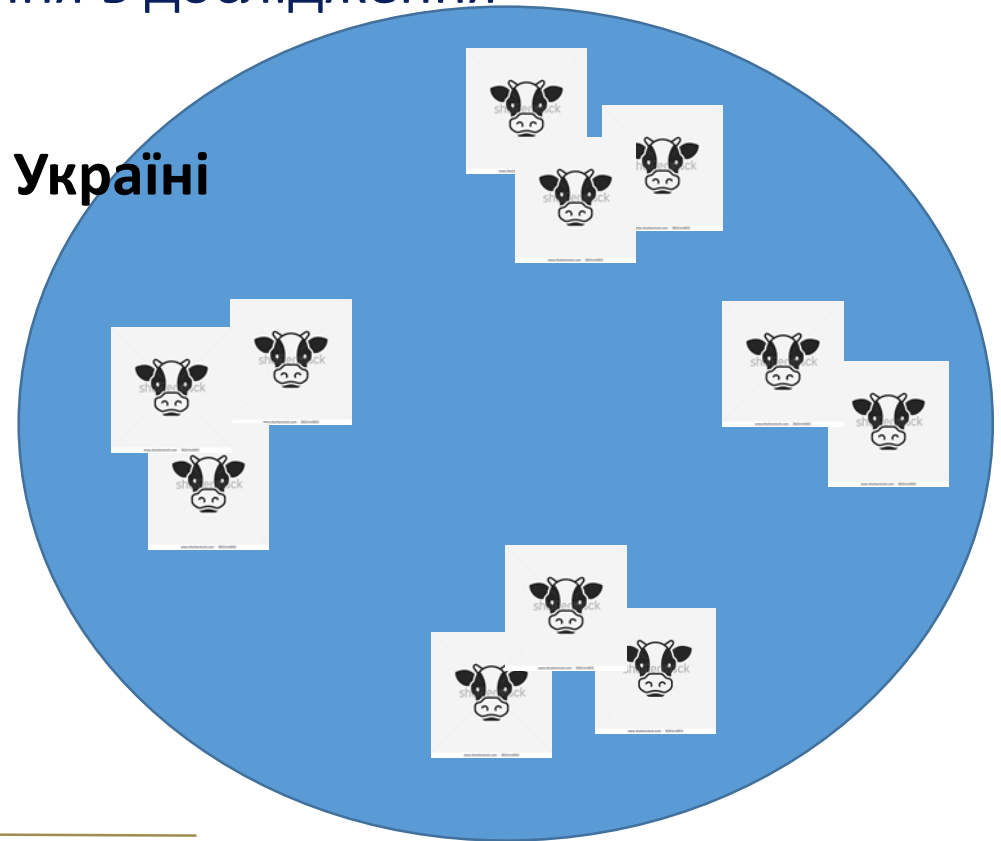
- **Мастити голштинської породи на фермах середніх розмірів в Україні**

Джерело популяції =

усі корови голштинської

породи на фермах середніх розмірів

у 4 областях, які охоплює проект



# Дослідження вибірки

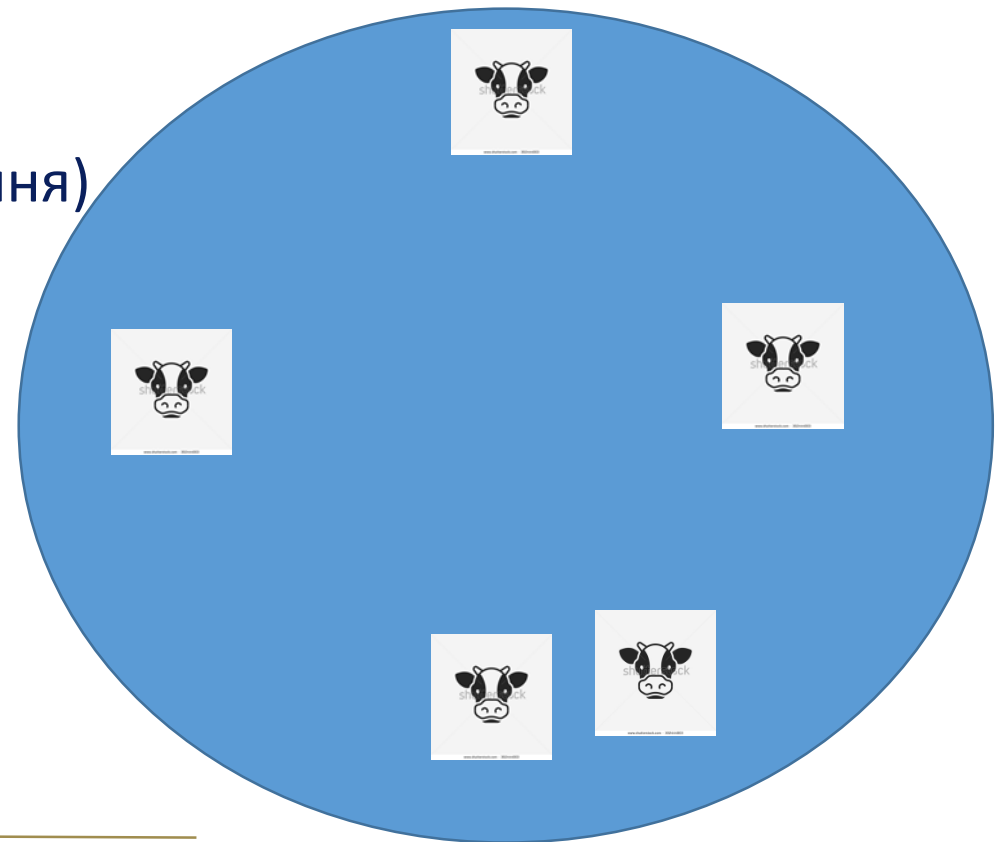
- Дослідження вибірки складається з **індивідуумів** (наприклад, тварини, стада, села), які увійдуть у дослідження
- До проведення дослідження дослідники мають визначити необхідний **розмір вибірки**

(тобто скільки індивідуумів необхідно включити у дослідження)

- **Мастити голштинської породи в Україні:**

Дослідження вибірки=

вбірка (статистично відповідна)  
голштинської породи у 4 областях,  
які охоплює проект



# Внутрішня валідність і зовнішня валідність

- **Внутрішня валідність дослідження:**

- Чи є результати дослідження (отримані з вибірки) валідними і для членів вихідної популяції (джерела популяції)
- Це свідчить про те, чи дійсно дослідження отримало "правильну" відповідь для джерела популяції
- Статистичні та епідеміологічні дослідження намагаються максимізувати ймовірність отримання правильної відповіді

- **Зовнішня валідність дослідження**

- Наскільки добре ці результати можуть бути узагальнені для цільової популяції
- Суб'єктивна оцінка наявності або відсутності широкої представленості цільової популяції у вихідній популяції
- Оцінка зовнішньої валідності набагато складніша, ніж оцінка внутрішньої валідності





# База вибірки, одиниці вибірки

- Перед тим, як обирається вибірка, члени вихідної популяції повинні бути визначені: база вибірки (список сіл, стад, тварин і т.д ...)



- Кожен член бази – одиниця вибірки:
  - окремі тварини,
  - стада, бійні, адміністративні регіони
- Превалентність (або інші індикатори) можуть бути обчислені по відношення до різних одиниць:
  - індивідуальна превалентність тварин = частка уражених тварин
  - превалентність стада = частка уражених стад



# База вибірки: обмеження

**а) Не завжди просто,**

- Список членів бази є **неповним**,
- Інформація стає **застарілою**,
- Сегменти бази **не прослідковуються**,
- Відсутність **співпраці** між деякими членами-категоріями бази,
- Прийнята процедура відбору **не є випадковою**,



Результати базуються на необ'єктивній оцінці



# База вибірки

**б) Бази не існує (табуни в кочових районах)**



# Стратегія відбору проб

- 1) Визначений відбір проб
- 2) Ймовірний відбір проб

## 1) Визначений відбір проб

- Вибір зразка залишається за дослідником
  - Конформна вибірка: легкодоступні одиниці вибірки,
  - Навмисна вибірка: характеристики вибірки (вік, стать, порода)



**Дуже малоймовірно, що вибірка буде дійсно представляти досліджувану популяцію**

- необ'єктивні результати



# Стратегія відбору проб

## 2) Ймовірний відбір проб

- Відбір зразка проводиться за допомогою неупередженого процесу тому ...

кожна одиниця вибірки в групі має ту ж ймовірність вибору  
(випадкова вибірка)

а) Проста випадкова вибірка

а) Систематичний відбір

а) стратифікована випадкова вибірка

а) кластерна вибірка



## а. Проста випадкова вибірка

- Визначення **баз** усіх тварин або одиниць вибірки:
- Використання таблиці випадкових чисел
- Комп'ютерні програми (SurveyToolbox, Episcopes, EpiInfo, Excel) або кишеньковий калькулятор виберуть випадкові вибірки,
- Опубліковані таблиці доступні:



# Таблиця випадкових чисел

|      |             |      |             |      |
|------|-------------|------|-------------|------|
| 7766 | 5509        | 2494 | <b>0122</b> | 9027 |
| 7481 | <b>2870</b> | 0407 | 4411        | 0892 |
| 0588 | 5444        | 6724 | <b>3822</b> | 8928 |
| 0579 | 9369        | 6921 | 1533        | 6138 |



# Проста випадкова вибірка

## Переваги

- Легко застосувати (якщо є повний список бази)
- Корисна для конкретних ситуацій, наприклад, відбираючи зразки 10% лабораторій підтверджуєш випадки сказу протягом останніх 5 років
- Може бути компонентом більш складних конструкцій

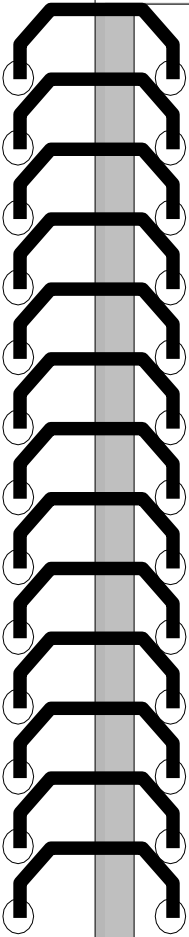
## Недоліки

- Потребує повної бази зразків
- Можливо важко реалізувати в полі
- Схильність до надмірної представленості або недопредставленості певних секцій бази вибірки





# Проста випадкова вибірка



|    |                 |    |                |
|----|-----------------|----|----------------|
| 1  | Albert D.       | 25 | Monique Q.     |
| 2  | Richard D.      | 26 | Régine D.      |
| 3  | Belle H.        | 27 | Lucille L.     |
| 4  | Raymond L.      | 28 | Jérémy W.      |
| 5  | Stéphane B.     | 29 | Gilles D.      |
| 6  | Albert T.       | 30 | Renaud S.      |
| 7  | Jean William V. | 31 | Pierre K.      |
| 8  | André D.        | 32 | Etienne M.     |
| 9  | Jeremy W.       | 33 | Marie M.       |
| 10 | Anthony Q.      | 34 | Gaétan Z.      |
| 11 | James B.        | 35 | Fidèle D.      |
| 12 | Denis G.        | 36 | Maria P.       |
| 13 | Amanda L.       | 37 | Anne Marie G.  |
| 14 | Jennifer L.     | 38 | Michel K.      |
| 15 | Philippe K.     | 39 | Gaston C.      |
| 16 | Eve F.          | 40 | Alain M.       |
| 17 | Priscilla O.    | 41 | Olivier P.     |
| 18 | Robert D.       | 42 | Geneviève M.   |
| 19 | Brian F.        | 43 | Berthe D.      |
| 20 | Hellène H.      | 44 | Jean Pierre P. |
| 21 | Isabelle R.     | 45 | Jacques B.     |
| 22 | Jean T.         | 46 | François P.    |
| 23 | Samanta D.      | 47 | Dominique M.   |
| 24 | Berthe L.       | 48 | Antoine C.     |

## b. Систематичний відбір

- Надає результати, аналогічні простій випадковій вибірці

Передбачає:

- Вибір одиниць вибірки через **рівні** проміжки часу,
- Тільки **перша тварина** повинна бути обрана випадковим чином.
- Застосування в польових умовах:

1 тварина із кожних 100 передбачає:

- 1) Випадковий вибір 1 тварини з перших 100 (s.r.s): n.63
- 2) Систематичний відбір: 63, 163, 263 etc...

- Виробничий контроль якості: вибір зразка на конвеєрній стрічці



# Систематичний відбір

## Переваги

- Спрощує вибір зразків: рандомізація потрібна тільки на старті
- Не вимагає знання загальної чисельності досліджуваної популяції

## Недоліки

- Якщо популяція не випадково представлена, але коливаються періодично, то оцінка може упереджено.

Наприклад, фермер відправив тварину на забій у понеділок, а проби на бойні відбираються тільки в четвер, тому тварина фермера не буде представлена в зразку



## с. Стратифікована вибірка

- Необхідно розділити досліджувану популяцію на **групи (страти)** відповідно до факторів, які, як очікується, можуть вплинути на результат (розмір стада, система управління, географічні райони),
- Долає обмеження простої випадкової вибірки, зокрема недопредставленість якоїсь групи,
- Кількість одиниць вибірки в кожній групі:

а. Проста випадкова вибірка

б. Пропорційний розподіл



# Стратифікована вибірка

- Пропорційний розподіл: кількість одиниць вибірки, обраних з кожної групи, пропорційна числу кожної групи :

наприклад, 500.000 молочної худоби поділити:

регіон «х»: 300.000 ВРХ

регіон «у»: 200.000 ВРХ

**Необхідний розмір вибірки 5%: 25.000**

регіон «х»:  $300.000 * 0.05 = 15.000$

регіон «у»:  $200.000 * 0.05 = 10.000$



# Стратифікована вибірка

## Переваги

- можна досягнути підвищення точності показників вибірки
- В межах груп показники є доступними

## Недоліки

- Знання елементів в популяції повинні бути відомі заздалегідь, щоб помістити у відповідні групи
- Декілька факторів можуть вплинути на результат, але не практично використовувати для стратифікації більше ніж один або два фактори
- Поганий вибір факторів стратифікації може привести до зниження точності



# Розмір вибірки



# Розмір вибірки

Залежить від:

## 1) Нестатистичні фактори

- Трудові ресурси, час, логістика,
- Ресурси,
- Наявність бази

## 2) Статистичні фактори

- Очікувана точність оцінки превалентності
- Очікувана превалентність хвороби





# Розмір вибірки

- **Бажана точність**

- Здатність оцінювача визначити за допомогою репрезентативної вибірки **дійсне значення** змінної для популяції,
- Це може бути виражено в термінах – межа **похибки оцінки**, яку можна допустити :
- Наприклад, абсолютна похибка  $\pm 2\%$  від оцінки 40% превалентності: 38-42%
- відносна похибка: 2% від 40%. Діапазон:  $40 \pm 0,8\%$



# Розмір вибірки

- **Очікувана превалентність**

- Парадоксально?

- Превалентність з однаковим ступенем точності :

очікуваний рівень превалентності ближче до 0% або 100%,  
потрібно менший розмір вибірки, ніж превалентність близько 50%,

Інформація не доступна: оцінка = поінформовані здогадки  
(приблизна оцінка)



# Розмір вибірки для оцінки **превалентності**

- **Що потрібно?**
  - Очікувана превалентність (**P**),
  - Бажана абсолютна точність (**d**: 1%, 5%, 10%)
  - Рівень достовірності ваших результатів (**C.I.**: 90%, 95%, 99%)
- Велика популяція:  $n = 1.96^2 P(1-P)/d^2$  де 1.96 відповідає 95% C.I.
- Маленька популяція: скоригована формула



# Розмір вибірки для оцінки превалентності

- недосконалі тести ( $Se, Sp < 100\%$ )

$$n = \frac{((1.96/d)^2 * ((Se * P) + (1-Sp)(1-P)) * ((1-Se*P)-(1-Sp)(1-P)))}{(Se+Sp-1)^2}$$

- опубліковані таблиці (*Cannon and Roe, 1982*)
- програмне забезпечення: Epiinfo, Stata etc...



# Розмір вибірки для виявлення захворювань

- $n = (1 - (1 - P)^{1/d}) \cdot ((N - d/2) + 1)$

**N** = розмір популяції (наприклад, стадо 200 тварин)

**d** = число очікуваних уражених тварин в популяції

Висококонтагіозні захворювання: використання мінімум серопревалентності (наприклад, принаймні 5% )

Не контагіозні захворювання: використовувати максимальну серопревалентність

**P** = ймовірність знаходження принаймі 1 випадок в зразку (наприклад, 0.95)



**n = 51** “Якщо серопревалентність 5%, то 51 тварину необхідно обрати, щоб виявити щонайменше одну серопозитивну тварину з ймовірністю 0.95”



# Активне спостереження на птахівництві за пташиним грипом



- Рішення **Європейської Комісії** від 13 квітня 2007 року про здійснення програм спостереження за пташиним грипом серед домашньої птиці та диких птахів у державах-членах



# Стратегія відбору проб

- Відбір проб повинен бути **стратифікованим** на всій території країни, беручи до уваги:

(а) Кількість господарств для відбору;

кількість має бути визначена таким чином, щоб забезпечити ідентифікацію щонайменше одного зараженого господарства, якщо превалентність заражених господарств становить  $\geq 5\%$  (95% інтервал - 99% для качок, гусей та індиків)



**Кількість господарств, з яких мають бути відібрані проби по кожній категорії птиці (крім господарств, які тримають індиків, качок і гусей)**

| Кількість господарств по видам птиці | Кількість господарств для відбору проб |
|--------------------------------------|----------------------------------------|
| До 34                                | Всі                                    |
| 35-50                                | 35                                     |
| 51-80                                | 42                                     |
| 81-250                               | 53                                     |
| >250                                 | 60                                     |





## Кількість господарств, які тримають індиків, качок і гусей, в яких повинні бути взяті проби

| Кількість господарств по видам птиці | Кількість господарств для відбору проб |
|--------------------------------------|----------------------------------------|
| До 46                                | Всі                                    |
| 47-60                                | 47                                     |
| 61-100                               | 59                                     |
| 101-350                              | 80                                     |
| >350                                 | 90                                     |



# Стратегія відбору проб

(b) кількість птахів, що відбираються з кожного господарства повинні бути визначені таким чином, щоб забезпечити 95% вірогідності ідентифікації щонайменше однієї позитивної птиці, якщо превалентність серопозитивних птахів становить  $\geq 30\%$ .



Зразки крові для серологічного дослідження повинні бути відібрані щонайменше від 5 до 10 птахів в кожному господарстві кожного виду птиці (крім качок, гусей, перепелів),

Якщо в господарстві є більше, ніж один сарай, то рекомендується відбирати не менше 5 особин з кожного сараю.





Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,  
Education and Research EAER  
**State Secretariat for Economic Affairs SECO**



Дякую!