



Віталій Кісіль



ОЦІНЮВАННЯ ЯКОСТІ Й НАДІЙНОСТІ ІНФОРМАЦІЇ ЩОДО ЗДОРОВ'Я, ЗІБРАНОЇ ЗА ДОПОМОГОЮ ІНТЕРНЕТ-ПАНЕЛЕЙ

Сьогодні дедалі більшої популярності набувають соціологічні опитування за допомогою Інтернету. Їх перевагами можна вважати швидкість збору даних та порівняну дешевизну. Однак у цього методу є й суттєві недоліки. Найбільший із них той, що не в усього населення є доступ до Інтернет. Наприклад, в Україні, за даними КМІС, станом на лютий 2011 року проникнення Інтернет серед дорослого (старшого 18 років) населення України в цілому сягало 30%, причому майже половину інтернет-користувачів становить молодь 18–29 років (46,6%)¹. Звісно, в США та інших країнах з розвинутою економікою й вищим рівнем життя Інтернетом користується більша частка мешканців, однак і в цих країнах проникнення Інтернет є неповним і нерівномірним².

Таким чином, з поля зору дослідників, які проводять опитування за допомогою Інтернет, просто випадає значна частина населення країни, тому не можна говорити про репрезентативність дослідження. Однак це зовсім не означає, що Інтернет-опитування поки що не придатні для використання у соціологічних і маркетингових опитуваннях. Доказом цього є надзвичайна популярність тематики онлайн-опитувань серед дослідників. Наприклад, на 65-ій щорічній конференції AAPOR, яка відбувалась 13–16 травня 2010 року в місті Чикаго, США, Інтернет-

опитування були темою багатьох доповідей. Дослідники приділяють багато уваги питанням оцінювання надійності та репрезентативності даних, отриманих за допомогою онлайн-опитувань. Для такого оцінювання найчастіше порівнюють демографічні характеристики респондентів Інтернет-опитувань та інших опитувань з репрезентативною вибіркою.

Одне з таких досліджень було спрямоване на порівняння відповідей респондентів, які брали участь у національному опитуванні в США з комбінованими методами проведення інтерв'ю за допомогою мережі Інтернет і пошти. Попередні результати цього дослідження були представлені на щорічній конференції AAPOR його організаторами EunjungShin, Timothy P. Johnson та Kumar Raob. Дослідження базувалось на даних 2008 року, зібраних у панелі здоров'я Gallup (Gallup Health panel survey). Панель Gallup становлять респонденти (кількістю 39 444 людини), випадково відібрані з усього населення США. Те, яким чином респондент заповнюватиме анкету, через Інтернет чи пошту, визначалось на основі його практики користування Інтернетом (що, у свою чергу, визначалось на етапі рекрутингу).

Результати показали, що в США респонденти, які брали участь в опитуванні через Інтернет, значно відрізняються від поштових респондентів за віком, статтю, расою, доходом домогосподарства, освітнім рівнем, урбанізацією та самоо-

¹ <http://inpoll.net/articles.php?id=2&lang=rus>

² <http://www.internetworldstats.com/stats.htm>

цінкою здоров'я. Загалом, онлайн-респонденти в цьому дослідженні були молодші, більш освічені, здоровіші й багатші, ніж респонденти, які відповідали поштою. Так само вони з більшою вірогідністю були одружені чи заміжні й мали власний дім. Порівняно з поштовими респондентами, серед Інтернет-респондентів було менше жінок, мешканців сільської місцевості та афроамериканців. Також дослідники з'ясували, що Інтернет-опитування загалом має нижчий респонс рейт (response rate) порівняно з опитуванням поштою, окрім вікової групи молодшої за 35 років, де респонс рейт серед онлайн-респондентів був вищим, ніж серед поштових. Попри це дослідники відзначили: є певні причини припускати, що, відповідаючи через Інтернет, респонденти надають якіснішу інформацію на відкриті запитання, однак це припущення ще потребує перевірки.

Отже, результати цього дослідження вказують на проблему невідповідності характеристик респондентів, які беруть участь в Інтернет-опитуваннях, з усім населенням США. З одного боку, це суттєва проблема, але перед дослідниками часто стоять завдання, які не потребують побудови репрезентативної вибірки, а спрямовані на дослідження думки певних специфічних груп та категорій населення. У таких випадках дослідники можуть звернутись до методу онлайн-опитувань, однак їм потрібно бути впевненими, що результати, зібрані через Інтернет, суттєво не відрізнятимуться від думки всіх представників цих груп чи категорій. На конференції AAPOR Джоном М. Бойлом (John M. Boyle) було представлено дослідження, направлене на оцінювання надійності ви-

користання Інтернет-панелей. Це дослідження під назвою «Здоров'я та Інтернет-панелі» було спрямоване на з'ясування доцільності використання методу онлайн-опитування для вивчення здоров'я населення.

Дослідження базувалось на двох телефонних опитуваннях, проведених у рамках дослідження астми та її лікування у США. Перше опитування мало національну випадкову вибірку розміром 2500 респондентів — хворих на астму віком від 12 років. Під час другого було опитано 1090 респондентів незалежно від того, хворіли вони на астму чи ні. Обидві вибірки побудовані за допомогою випадкового набору номера (RDD). Респондентів запитували, окрім питань про їхнє здоров'я, ще й про те, чи брали вони участь в Інтернет-опитуваннях і як часто. Таким чином, дослідники отримали змогу порівняти демографічні характеристики й відповіді щодо хвороби на астму респондентів, які не беруть

участі в Інтернет-панелях, з тими, які принаймні раз на місяць проходять онлайн-опитування.

Результати дослідження показали, що учасники Інтернет-панелей відрізняються від неучасників за демографічними характеристиками (вік, дохід, раса й етнічність, сімейний стан, зайнятість і тип домогосподарства) та деякими аспектами здоров'я в обох вибірках. Однак опитування пацієнтів, хворих на астму, показало, що, попри різницю в демографічних характеристиках, відповіді респондентів — учасників онлайн-панелей щодо хвороби не відрізняються суттєво від відповідей респондентів, які не беруть участі в Інтернет-опитуваннях. Відповіді обох категорій респондентів були майже однаковими щодо: загального стану здоров'я, частоти й тяжкості симптомів, тягаря, викликаного хворобою, госпіталізацією та викликом швидкої допомоги, а також щодо вживання медикаментів.

Таблиця 1. Поділ респондентів обох опитувань за статтю та віком

	Репрезентативна вибірка		Онлайн-вибірка	
	кількість, осіб	частка, %	кількість, осіб	частка, %
Чоловік	459	45,0	546	79,6
Жінка	561	55,0	140	20,4
Всього	1020	100,0	686	100,0
18–29	227	22,2	480	70,0
30–39	178	17,4	143	20,8
40–49	196	19,2	45	6,6
50–59	141	13,9	9	1,3
60–69	149	14,6	7	1,0
70 +	129	12,7	2	0,3
Всього	1020	100,0	686	100,0

Таблиця 2. Самооцінка стану здоров'я респондентами з репрезентативної та онлайн-вибірки

Самооцінка здоров'я	Репрезентативна вибірка		Онлайн-вибірка (не зважена)		Онлайн-вибірка (зважена за статтю та віком)	
	кількість, осіб	частка, %	кількість, осіб	частка, %	кількість, осіб	частка, %
Дуже погане	31	3,0	4	0,6**	7	1,0**
Погане	172	16,9	19	2,8**	17	2,5**
Посереднє (ані хороше, ані погане)	469	46,0	190	27,7**	363	52,9**
Хороше	309	30,3	380	55,4**	262	38,3**
Дуже хороше	31	3,1	88	12,8**	31	4,5
Важко сказати	5	0,5	3	0,4	5	0,7
Немає відповіді	3	0,2	2	0,3	1	0,1
Всього	1020	100,0	686	100,0	686	100,0

*Різниця значуща на рівні $p = 0,05$ **Різниця значуща на рівні $p = 0,01$ **Таблиця 3.** Поділ респондентів обох опитувань віком 18–29 років за статтю та типом населеного пункту

	Репрезентативна вибірка		Онлайн-вибірка	
	кількість, осіб	частка, %	кількість, осіб	частка, %
Чоловік	114	50,5	392	81,7
Жінка	112	49,5	88	18,3
Всього	227	100,0	480	100,0
Сільське населення	63	27,8	35	7,3
Міське населення	164	72,2	445	92,7
Всього	227	100,0	480	100,0

Отже, можна зробити висновок, що використання Інтернет-панелей для опитування певних специфічних груп респондентів (наприклад, хворих на астму) дає досить якісні результати там, де точні результати й довірчі інтервали є не настільки критично необхідними. Зважаючи ж

на зростання Інтернет-покриття, якість онлайн-опитувань лише поліпшуватиметься.

З огляду на описані вище дослідження можна припустити, що, корегуючи співвідношення цих груп у онлайн-вибірці, можна наблизити її до репрезентативної для всієї країни. З метою перевірки цієї гіпотези Київським міжнародним

інститутом соціології та компанією InPoll здійснено два паралельні опитування, одне з яких було з репрезентативною вибіркою для дорослого населення України, старшого 18 років, а інше проводилось серед респондентів інтернет-панелі InPoll. Унікальністю даного дослідження є те, що обидва опитування проводились одночасно і містили однакові запитання, тому в дослідників була можливість прямо порівняти їхні результати.

Як уже зазначалося вище, одне опитування було проведене з репрезентативною вибіркою, тому розподіл респондентів за віком і статтю повністю відповідає такому для населення, старшого 18 років, по всій Україні, однак ці характеристики у вибірці для Інтернет-опитування значно відрізнялись (табл. 1). Як видно, в онлайн-вибірці значно переважає частка чоловіків та молодих людей віком 18–29 років, тоді як частка респондентів, старших 40 років суттєво нижча, ніж у репрезентативній вибірці.

Після зіставлення відповідей респондентів обох опитувань на питання щодо їхнього оцінювання свого здоров'я виявилось, що респонденти онлайн-опитування в цілому, позитивніше оцінюють стан свого здоров'я (табл. 2).

Зважування за статтю та віком дещо змістило відповіді респондентів онлайн-опитування у бік відповідей респондентів із репрезентативної вибірки, але різниця у розподілах все одно значуща. Поясненням цьому може бути незначна наповненість вікових категорій людей, старших 40 років в онлайн вибірці: хай би яким чином ми їх не зважували, наприклад, сім осіб віком 60–69 років не можуть вважатися репрезентативною групою для всієї України. З іншого боку, можна

припустити, що відповіді всередині наповненої групи з онлайн-опитування можуть бути репрезентативними. Тому для перевірки цієї гіпотези можна порівняти відповіді респондентів віком 18–29 років, оскільки ця група є найбільш наповненою в опитуванні через Інтернет.

Як і у випадку всіх респондентів, серед респондентів віком 18–29 років в обох вибірках існують суттєві розбіжності за статевим складом (табл. 3).

Через малу наповненість категорії сільського населення у вибірці онлайн-опитування InPoll було вирішено проводити зважування лише за статтю. Виявилось, що навіть незважені відповіді респондентів онлайн-опитування InPoll віком 18–29 років майже повністю збігаються з відповідями респондентів із репрезентативної вибірки. Лише кількість тих, хто оцінює своє здоров'я як «дуже хороше», вище серед онлайн-респондентів InPoll. Зважування за статтю допомогло вирівняти цю різницю і в результаті відповіді респондентів обох опитувань віком 18–29 років статистично не відрізняються (табл. 4).

Зважування за статтю також допомогло наблизити відповіді респондентів обох вибірок віком 30–39 років (ця група теж є достатньо наповненою в онлайн-вибірці InPoll). Якщо до зважування відповіді дещо відрізнялись, то після нього статистично значущої різниці не помічено (табл. 5).

Отже, видно, що, зважаючи на статть, можна отримати репрезентативні або принаймні такі, які дуже близькі до репрезентативних, результати опитування, проведеного в Інтернеті. Однак для цього повинна бути виконана умова достатньої наповненості всіх основних

Таблиця 4. Самооцінка стану здоров'я респондентами віком 18–29 років з репрезентативної та онлайн-вибірки

Самооцінка здоров'я	Репрезентативна вибірка		Онлайн-вибірка (не зважена)		Онлайн-вибірка (зважена за статтю)	
	кількість, осіб	частка, %	кількість, осіб	частка, %	кількість, осіб	частка, %
Дуже погане	1	0,5	1	0,2	1	0,1
Погане	8	3,3	11	2,3	19	4,0
Посереднє (ані хороше, ані погане)	66	29,0	110	22,9	145	30,2
Хороше	129	56,7	281	58,5	255	53,1
Дуже хороше	23	10,0	75	15,6*	59	12,3
Важко сказати	0	0,0	1	0,2	1	0,1
Немає відповіді	1	0,5	1	0,2	1	0,1
Всього	227	100,0	480	100,0	480	100,0

Таблиця 5. Самооцінка стану здоров'я респондентами віком 30–39 років з репрезентативної та онлайн-вибірки

Самооцінка здоров'я	Репрезентативна вибірка		Онлайн-вибірка (не зважена)		Онлайн-вибірка (зважена за статтю)	
	кількість, осіб	частка, %	кількість, осіб	частка, %	кількість, осіб	частка, %
Дуже погане	2	1,3	2	1,4	4	2,5
Погане	6	3,5	6	4,2	6	4,1
Посереднє (ані хороше, ані погане)	90	50,5	46	32,2**	60	41,8
Хороше	76	43,0	77	53,8	64	45,0
Дуже хороше	2	1,2	10	7,0**	6	4,2
Важко сказати	1	0,5	1	0,7	3	2,0
Немає відповіді	0	0,0	1	0,7	1	0,4
Всього	178	100,0	143	100,0	143	100,0

*Різниця значуща на рівні $p = 0,05$

**Різниця значуща на рівні $p = 0,01$



соціально-демографічних категорій у онлайн-вибірці. Якщо ці категорії будуть наповнені, то ми можемо отримати такі ж дані, як і при проведенні звичайного репрезентативного опитування face-to-face. Звісно, зараз існує проблема досягнення через Інтернет старших респондентів і мешканців сіл, але видно, що вже зараз онлайн-панель InPoll може забезпечити досить репрезентативні результати у разі опитування респондентів віком 18–39 років. Тому в Україні вже зараз можна користуватись методом онлайн-опитування для проведення соціологічних і маркетингових досліджень цієї вікової категорії. Проблему ж важкої досяжності старших людей можна вирішити, надаючи відібраним респондентам доступ до Інтернету або ж просто опитувати їх за допомогою face-to-face, що всеодно буде швидше й дешевше, адже обсяг вибірки для особистого інтерв'ю значно зменшиться порівняно зі стандартними репрезентативними опитуваннями. ■