

# Konferencja SONDAŻ POLSKI. METODA, ETYKA, MEDIA



Kto wypełnia te ankiety przez internet?  
I co z tą reprezentatywnością?



Michał Kot  
Dyrektor Badań w IIBR

# Badania bez ankietera

Na początku był ankieter, który:

- wyjaśniał trudne pytania
- podpowiadał, powtarzał
- zatrzymał respondenta dłużej (70 minut)

... a kiedy ankietera zabrakło, nikt:

- nie wyjaśnił
- nie powtórzył
- nie zatrzymał



... dlatego w pierwszych badaniach online wiele było:

- przerwanych ankiet
- nielogicznych odpowiedzi
- problemów z kontrolą prób

# W badaniach spotkały się dwa światy



Klasyczny



Współczesny

\* Zdjęcia pochodzą z Wikipedii



# Trzeba było próbować i nauczyć się to robić na nowo **ONLINE!**

- Jak dobrać próbkę?
- Jak zapraszać do badania?
- Jak budować kwestionariusze (struktura, czas)?
- Jak projektować pytania kwestionariuszowe?
- Jak utrzymać uwagę w kwestionariuszu (wszystkich)?
- Jak porządkować zebrane dane?
- Jak ważyć dane?
- Jak analizować?



Wysoka jakość danych!

# Skąd się biorą respondenci online?

## RTS

## Panel badawczy

Wszyscy internauci

Użytkownicy stron\*

Użytkownicy stron w czasie realizacji badania

Wszyscy internauci

Użytkownicy stron\*

Osoby, które wzięły udział w jakimkolwiek badaniu z rekrutacją do Panelu



Osoby, które zapisały się do Panelu

Osoby, które **w momencie emisji** zaproszenia miały czas i chęć wziąć w nim udział.  
**JEDNORAZOWO**

Osoby, które **w czasie realizacji badania** miały czas i chęć wziąć w nim udział.  
**MOŻNA WRACAĆ. ZAPROSZENIA**

*\* najczęściej są to grupy witryn internetowych obsługiwanych kompleksowo przez wydawców, powierzchniowo płatne za emisję*

## Kto wypełnia – różnice w profilu

|                                       | RTS  | Panel   |
|---------------------------------------|--|---|
| <b>Płeć</b>                           |  | więcej kobiet   |
| <b>Wiek</b>                           | więcej młodych   |   |
| <b>Wykształcenie</b>                  |  | więcej lepiej wykształconych  |
| <b>Miejsce zamieszkania</b>           |  | więcej osób z dużych miast  |
| <b>Sytuacja zawodowa</b>              |  | więcej osób ze stałą pracą  |
| <b>Dochody</b>                        | więcej gorzej zarabiających  |   |
| <b>Wielkość gospodarstwa domowego</b> | więcej osób z dużych gospodarstw   |   |
| <b>Staż w internecie</b>              |  | dłuższy   |
| <b>Częstotliwość korzystania</b>      | (ale też heavy users)  | więcej heavy users  |
| <b>Kompetencje komputerowe</b>        |  | wyższe<br> |

# Kto wypełnia – różnice w wynikach

|   | RTS      | Panel   |
|---|----------|---------|
| Znajomość marek   |          | wyższa  |
| Używanie produktów  |          | wyższe  |
| Intencja zakupu produktów                                     |          | wyższa  |
| Zapamiętanie reklam marki                                     | wyższe   |         |
| Unikanie odpowiedzi,<br>porzucanie ankiety                    | częstsze |         |
| Liczba wskazań w długich<br>pytaniach wielokrotnego<br>wyboru |          | większa |



# Kiedy stosować

## RTS

- szerokie próby celowe, także reprezentatywne
- krótsze, prostsze ankiety
- grupy docelowe definiowane behawioralnie lub geolokalizacyjnie (szerokie oraz wąskie)
- specyficzne grupy docelowe (targetowane pod kątem zachowań w internecie, kontaktu ze stronami, reklamami)
- łączenie z „twardymi danymi”

## Panel

- próby celowe
- dłuższe, bardziej złożone ankiety
- wąskie grupy docelowe
- powroty do respondenta (profil demograficzny, styl życia, używanie produktów, używanie specyficznej kategorii itd.)
- testy poufnych materiałów
- komunikacja w obie strony

**Narzędzia te mają różne zastosowania ze względu na specyfikę!**

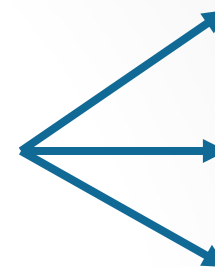


# Z czym się mierzymy w badaniach online



# Najczęstsze obawy wobec badań online

Problem doboru próby



Niska stopa zwrotu,  
specyfika odpowiadających



Brak kontroli ankietera – „klikacze” i „psotnicy”



# Jak dobierać próbę online

Penetracja internetu w  
wybranych grupach

- **Populacja**

- Polacy ≠ internauci (jeszcze 😊)

→ nie powinno się badać grup o niskiej penetracji internetu  
(ALE większość typowych TG możemy spokojnie badać)

- **Metody doboru**

- brak „spisu internautów”, niejednoznaczność identyfikacji

→ identyfikacja cookie, weryfikacja struktury

95%  
osób z wyższym  
wykształceniem

95%  
uczniów  
i studentów

85%  
osób w wieku  
15-45 lat

75%  
osób mających  
stałą pracę

\*NOTA: Cookie nie jest jednoznacznym id użytkownika (różne przeglądarki), więcej przeglądarek, różne osoby przy komputerze

# Od czego tak naprawdę zależy trafność badań?

## Jak poznać „wynik prawdziwy”?

- ~~badanie wyczerpujące (całej populacji)~~
- badanie „próby reprezentatywnej”

## Co to jest próba reprezentatywna?

- losowa i w pełni zrealizowana (?)
- kwotowa/warstwowa, „reprezentatywna pod względem” (?)
- taka, w której osoby dostępne do badania nie różnią się od niedostępnych pod względem cech związanych z badanymi zjawiskami (!)



# RTS – co zrobić, nie znając profilu użytkownika

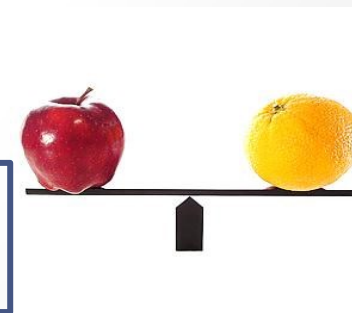
## Co wiemy o użytkowniku (pliki cookies\*)

- informacje techniczne i behawioralne
  - spis cookies (*gemiusTraffic*), losowanie
  - losowy-warstwowy dobór cookies do badania (populacja cookies jest heterogeniczna – heavy i light users )
  - regulowana emisja zaproszeń (wiek, aktywności cookies, cykl zaproszeń, capping)
  - zakładanie kwot minimalnych na niedoszacowane grupy
  - emitowanie zaproszeń na stronach o łącznym zasięgu powyżej 80%



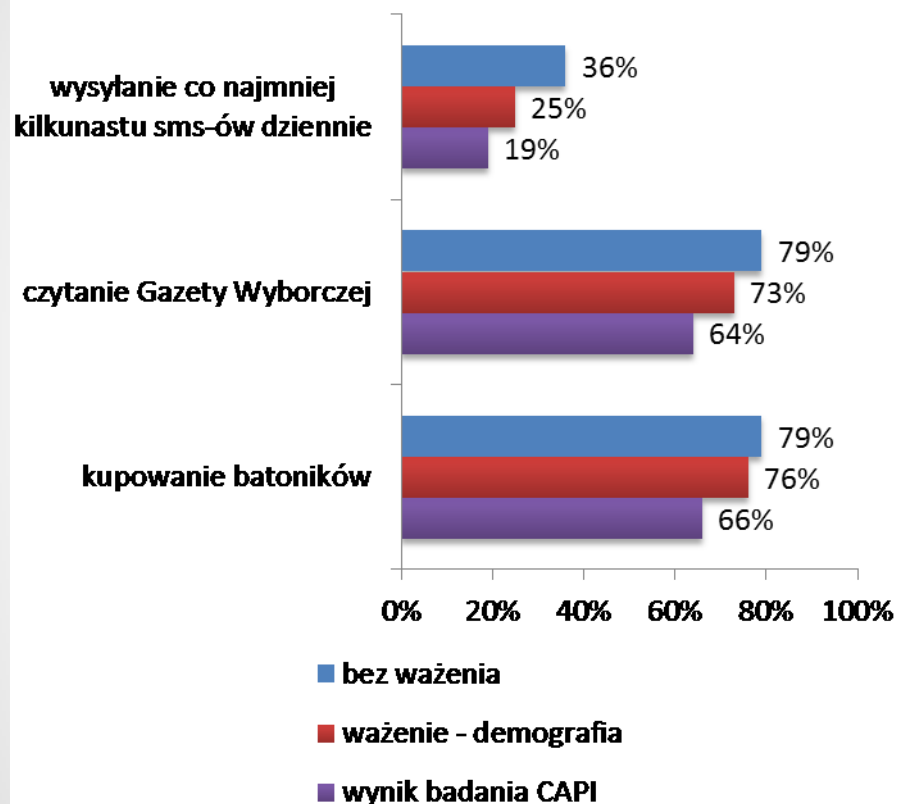
## Jak dobierać/korygować próbę, aby osiągnąć wysoką jakość wyników

- kluczowe zmienne wymagające kontroli
  - zmienne soc-dem
  - zmienne specjalne, behawioralne (ok. 10 zmiennych dopasowanych do typów badań)



# Korekta próby – demografia

## GRUPA CELOWA: POLACY 15-49 LAT

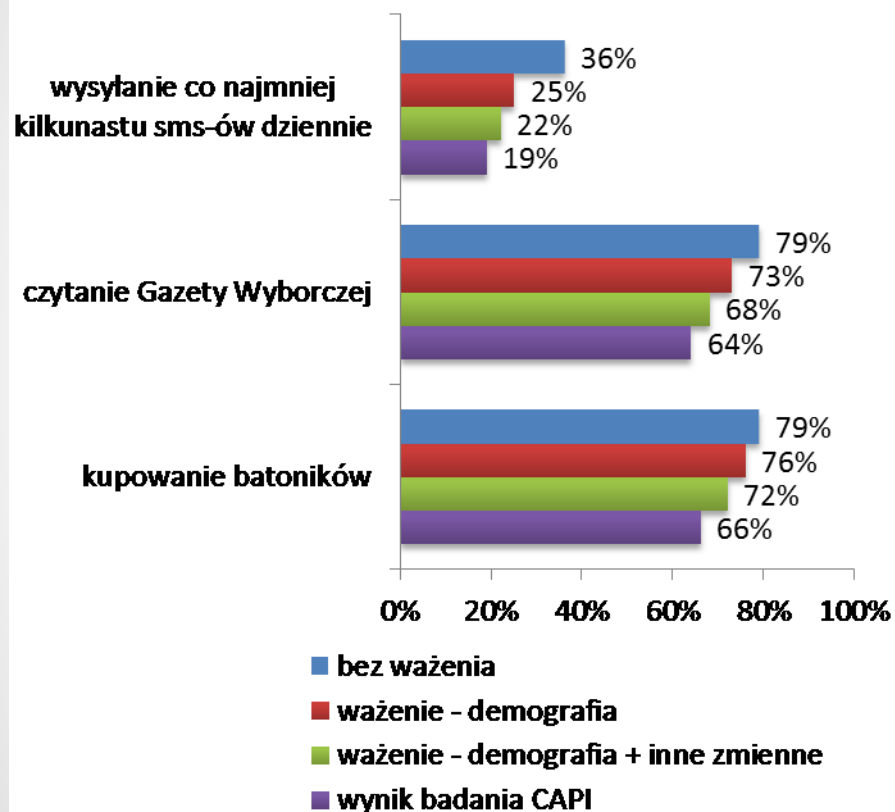


Kontrola „standardowych” zmiennych soc-dem przybliży dane do referencyjnych, ale i tak pozostaje spora różnica 😞

Źródło: badania w ramach projektu „Innowacyjne metody poprawy jakości badań prowadzonych przez internet”, maj 2010

# Korekta próby – wyjście poza demografię

## GRUPA CELOWA: POLACY 15-49 LAT



Nadzorowanie dodatkowych zmiennych kontrolnych przybliży dane do wyniku referencyjnego

Stosowanie w zależności od badanych grup różnych zmiennych dla zjawisk:

- **związanych z internetem**
- **niezwiązanych z internetem – populacja internautów**
- **niezwiązanych z internetem – grupy celowe szersze niż grupa internautów**

Źródło: badania w ramach projektu „Innowacyjne metody poprawy jakości badań prowadzonych przez internet”, maj 2010

# Stopa zwrotu i specyfika odpowiadających

## Response Rate

(%) RR = kilka (RTS) do kilkudziesięciu (Panel)

- atrakcyjne i nieinwazyjne zaproszenia
- optymalna budowa kwestionariusza
- konstrukcja i format pytań w kwestionariuszu
- dobór większej grupy potencjalnych respondentów



## Specyfika odpowiadających na ankiety

- Niektórzy respondenci mają mniej cierpliwości niż inni
  - uatrakcyjnianie ankiety (np. gamification)
  - kontrola struktury próby (technologia)





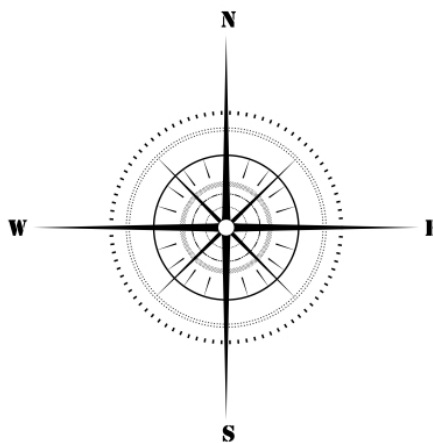
# A co z „Klikaczami” i „Psotnikami”?

## Kwestia „podejrzanych” danych (kilka %)

- Nieuważne/szybkie odpowiedzi („Klikacze”)
  - czytelne i jednoznaczne formułowanie pytań
  - atrakcyjność pytań
  - weryfikacja czasu wypełniania ankiety
- Manipulacja odpowiedziami („Psotnicy”)
  - weryfikacja spójności logicznej ankiety (pytania kontrolne)
  - kilka % zapasu w próbie, weryfikacja odpowiedzi, usuwanie nierzetelnych wpisów z bazy



# Jak pytać o standard badań online: panel i RTS



# Jak pytać o dobry standard badań online

1. Dobór próby do badania.
2. Sposób kontroli i korekty próby (struktura jakich zmiennych jest kontrolowana, w jakich przypadkach, do jakich danych referencyjnych, ew. też które rozkłady są łączne, a które brzegowe; sposoby korekty, np. próby kwotowe / ważenie / połączenie obu) wraz z uzasadnieniem wyboru takiego sposobu.
3. Karencja (jak często respondent może dostać ankietę).
4. Sposób kontroli spójności logicznej danych / czyszczenie danych.
5. Do jakich przeglądarek/wersji przeglądarek jest dostosowane działanie kwestionariuszy?
6. Fakt i sposób wynagradzania respondentów – czy respondent może łatwo przewidzieć, za jakie odpowiedzi dostanie wyższą nagrodę (formułowanie pytań screenujących)?
7. Sposób zapraszania respondentów do udziału w badaniu (aktywny vs pasywny).
8. W jaki sposób identyfikowani są respondenci?

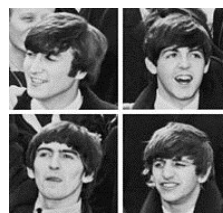
## Specyficzne do RTS

1. W jakim formacie wyświetlane są zaproszenia do badania (pop-up, pop-under, top-layer)?
2. Jaki jest łączny miesięczny zasięg standardowo wykorzystywanej powierzchni emisyjnej?
3. Na jakich powierzchniach emitowane są kwestionariusze z zaproszeniem do badania?
4. Czy jest możliwe targetowanie emisji zaproszeń do badania?

# Jak pytać o dobry panel online

1. Czy to jest własny panel?
2. Ilu respondentów jest zarejestrowanych w panelu?
3. Ile wynosi wskaźnik rotacji?
4. W jaki sposób respondenci są zapraszani do udziału w badaniu? Czy można zapisać się samemu?
5. Jak wygląda proces rejestracji do panelu?
6. Czy firma posiada procedury potwierdzające zgodność danych nowych panelistów oraz możliwość identyfikacji osób, które podają dane niezgodne ze stanem faktycznym?
7. Jakiego rodzaju dane o respondentach zawiera profil uczestnika panelu?
8. Jak często dane respondenta są aktualizowane?
9. Jak wygląda system wynagradzania panelistów?
10. Jak często ten sam uczestnik otrzymuje ankietę internetową w danym okresie?
11. Czy zachowywane są informacje na temat historii uczestnictwa panelistów w badaniach?
12. Czy w panelu obowiązuje polityka prywatności? Jeśli tak, jak ona wygląda?
13. Czy panel działa w zgodzie z wszystkimi regionalnymi, narodowymi i lokalnymi prawami respektowania prywatności, ochrony danych, ochrony dzieci itp. Czy używane są UE Safety Harbour i COPPA? Jakie inne branżowe standardy badań są wprowadzone, np. ICC/ESOMAR Międzynarodowy Kodeks Badań Rynku i Badań Społecznych, wytyczne CASRO itp.?
14. Jakie środki bezpieczeństwa/ systemy ochrony danych posiada firma?
15. Czy panel jest wykorzystywany do innych celów niż badania marketingowe?

# Klasyki się przyglądają



Młodzi choć nie stworzyli jeszcze swojego najlepszego albumu, to nagrali już kilka bardzo dobrych płyt!

\* Zdjęcia pochodzą z Wikipedii



Konferencja  
SONDAŻ POLSKI. METODA, ETYKA, MEDIA



Dziękuję

Michał Kot, IIBR  
[michal.kot@iibr.pl](mailto:michal.kot@iibr.pl)  
608 522 318

# IIBR - Najwyższa jakość badań online

## Projekty rozwojowe

Dofinansowanie z **Narodowego Centrum Badań i Rozwoju** w ramach programów:

1

### Innowacyjne metody poprawy jakości badań prowadzonych przez internet

- **Wyższa jakość prób** badawczych
- **Identyfikacja populacji do badań online**, tak aby uzyskiwać wiarygodne wyniki
- **Porównanie badań** online vs. offline



Realizacja 11'2009 – 11'2010,  
Budżet: 760 000 zł

2

### Innowacyjne techniki i narzędzia badań online minimalizujące obciążenie użytkowników

- **Krótszy czas przygotowania** badań bez straty jakości
- **Wyższa jakość danych** dzięki większemu zaangażowaniu respondenta w badanie
- **Bardziej efektywna dotarcie** do wąskich grup celowych bez straty jakości



Realizacja 02'2012 – 09'2013.  
Współpraca z SGH.  
Budżet: 1 882 000 zł

