

Maksymilian Gajda, Małgorzata Kowalska

## THE ROLE OF AN ELECTRONIC QUESTIONNAIRES IN EPIDEMIOLOGIC STUDIES AIMED AT CANCER PREVENTION

### MOŻLIWOŚĆ STOSOWANIA ELEKTRONICZNYCH KWESTIONARIUSZY W BADANIACH EPIDEMIOLOGICZNYCH UKIERUNKOWANYCH NA PROFILAKTYKĘ CHOROÓB NOWOTWOROWYCH

Medical University of Silesia, Medical Faculty in Katowice, Department of Epidemiology, Katowice, Poland  
Śląski Uniwersytet Medyczny w Katowicach, Wydział Lekarski w Katowicach, Katedra i Zakład Epidemiologii

#### ABSTRACT

The growing number of people diagnosed with cancer is an important issue in public health. Evaluation of the population's level of cancer-related knowledge is mandatory for planning effective prevention or intervention programs. Until recently, traditional methodology, using paper questionnaire or telephone-based interviews had dominated previous cross sectional studies related to this issue. Currently, the use of new research methods based on information technology solutions has been more frequent. Therefore, the aim of this paper is to review published data available in the PubMed database towards the identification of strengths and weaknesses of cross-sectional studies conducted via Internet, a type of information technology solution. This review allowed the authors to conclude that web-based questionnaire studies could be an effective solution and may possibly replace traditionally conducted interviews. Likewise, it evidently appears that an online mode is more adequate for querying young people. However, there is lack of consensus regarding the cost-effectiveness of the application of Web technology as a research tool for epidemiological questionnaire-based studies. Data security was revealed to be an important aspect of conducting Internet-based questionnaire studies. Further efforts are aimed at standardizing this type of researches to assess their strengths and weaknesses leading to more widespread use.

**Key words:** *questionnaire studies, Internet, prevention, cancer*

#### STRESZCZENIE

Rosnąca liczba osób z rozpoznaniem choroby nowotworowej jest istotnym problemem zdrowia publicznego. Poznanie stanu wiedzy populacji w tym zakresie wydaje się być ważne dla planowania skuteczniejszych programów profilaktycznych czy interwencyjnych. Do niedawna w badaniach przekrojowych dominowała tradycyjna forma, wykorzystująca kwestionariusze papierowe lub wywiady telefoniczne. Obecnie coraz częściej wykorzystuje się nowe metody badawcze oparte o rozwiązania informatyczne (głównie sieć Internet). Celem opracowania była identyfikacja mocnych i słabych stron badań epidemiologicznych prowadzonych z wykorzystaniem sieci Internet w oparciu o przegląd dostępnego w bazie PubMed piśmiennictwa. Zebrane dane literaturowe pozwalają wnioskować, iż badania kwestionariuszowe realizowane z zastosowaniem sieci WWW są rozwiązaniem skutecznym i mogą zastąpić tradycyjne formy zbierania wywiadów, jednakże raczej w odniesieniu do osób młodych. Ponadto ujawniono brak spójnego stanowiska na temat możliwości redukcji kosztów badań prowadzonych z wykorzystaniem technologii *Web* w porównaniu do technik tradycyjnych. Ważną kwestią w przypadku badań kwestionariuszowych prowadzonych drogą *on-line* jest konieczność zapewnienia bezpieczeństwa danych. Niezbędne są dalsze wysiłki zmierzające do standaryzacji stosowanych narzędzi wraz z oceną silnych i słabych stron prowadzonego badania.

**Słowa kluczowe:** *badania kwestionariuszowe, Internet, profilaktyka, choroby nowotworowe*

## INTRODUCTION

The growing number of people diagnosed with cancer is an important public health issue. Cancer related group of diseases is the second most common cause of death in Poland (1). There have also been a dynamic development in the IT sector (i.e internet) in recent years which parallel the growing evidence of cancer diagnoses in Poland. Current estimates puts the number of households in Poland with internet access at 75 percent (2). Evaluating the level of cancer related knowledge among the population in Poland seems to be mandatory in order to have an effective planning related to prevention and intervention. Review of published data indicates that cross-sectional studies are the most common method and effective means of assessing the level of knowledge regarding cancer prevention. Case-control as well as cohort studies have been conducted using interviews and self-tracking questionnaire according to the traditional way; with the aim of at gathering information regarding potential risk factors (3, 5). Newest research methods, which involve computer facilities (mainly the Internet) may help facilitate the effectiveness of conducting questionnaire-based studies (4-7). Analyzing available literature indicates that in Poland there has not been almost any research conducted on-line with the use information technology solutions (i.e. internet), on the assessment of the population's general awareness regarding cancer prevention (6). Given these circumstances, it seems important to define the role and determine the most appropriate use of the Internet for conducting the on-line questionnaire-based epidemiological studies.

## THE AIM OF STUDY

The aim of this study is to evaluate the possibility of conducting questionnaire-based epidemiological studies via Internet, an information technology solution. The specific aims in this review include methods of recruiting respondents, principles of constructing the questionnaire and its distribution, optimization of cross-sectional studies' costs.

## MATERIAL AND METHODS

To achieve the aim of this paper, reviews of published data available in PubMed database were conducted using the following key-words and phrases: *cancer; prevention; promotion; survey; knowledge; awareness; web-based*. We analyzed full-text articles published in the years 1997-2015.

## WSTĘP

Rosnącą liczbą osób z rozpoznaniem choroby nowotworowej jest istotnym problemem zdrowia publicznego. Schorzenia te stanowią w Polsce drugą przyczynę zgonów (1). Równocześnie, w ostatnich latach nastąpił dynamiczny rozwój sektora informatycznego, w tym ogólnosiwiatowej sieci Internet. W Polsce aktualnie ponad 75% gospodarstw domowych posiada dostęp do jej zasobów (2). Poznanie stanu wiedzy na temat chorób nowotworowych wydaje się istotne dla właściwego planowania skuteczniejszych programów profilaktycznych czy interwencyjnych. Przegląd aktualnego piśmiennictwa wskazuje, że najczęściej stosowaną metodą poznawania i oceny stanu wiedzy na temat profilaktyki chorób, w tym chorób nowotworowych, są badania przekrojowe. W epidemiologicznych badaniach kliniczno-kontrolnych lub kohortowych powszechnie stosowane są także wywiady lub kwestionariusze samozwrotne, których zadaniem jest m.in. zebranie informacji na temat potencjalnych czynników ryzyka choroby (3, 5). Nowe metody badawcze wykorzystujące jako narzędzie rozwiązania informatyczne (głównie sieć Internet) mogą stanowić skuteczną pomoc w realizacji badań kwestionariuszowych (4-7). Niestety, analiza dostępnego piśmiennictwa wskazuje, że praktycznie brak jest w naszym kraju badań dotyczących oceny ogólnej świadomości populacji w zakresie zapobiegania chorobom nowotworowym, realizowanych z wykorzystaniem możliwości on-line (6). Biorąc pod uwagę powyższe przesłanki, istotne wydaje się określenie roli i ustalenie najbardziej właściwych zastosowań Internetu dla prowadzenia epidemiologicznych badań kwestionariuszowych drogą *on-line*.

## CEL PRACY

Celem opracowania była identyfikacja możliwości wykorzystania sieci Internet do prowadzenia kwestionariuszowych badań epidemiologicznych. Wśród celów szczegółowych znalazły się takie aspekty jak: możliwość rekrutacji badanych, zasady konstrukcji kwestionariusza badawczego i jego dystrybucja, optymalizacja kosztów realizacji badań przekrojowych.

## MATERIAŁ I METODY

Dla realizacji celu pracy dokonano przeglądu piśmiennictwa dostępnego w zasobach bazy PubMed z wykorzystaniem następujących słów kluczowych: *cancer; prevention; promotion; survey; knowledge; awareness; web-based*. Analizie poddano pełnotekstowe artykuły opublikowane w latach 1997 – 2015.

## DISCUSSION

**Preliminary information**

Until recently, most of the cross-sectional studies had been conducted according to the traditional way by qualified interviewers using paper based questionnaire or telephone-based interviews performed. The advantages and disadvantages of using this modalities in medical science have been extensively discussed in epidemiology (3, 5). A major drawback to this type of traditional methodology is an increase in the number of refusals of the survey's potential participants. This trend leads to difficulties relating to the numbers of the study sample (5, 8). The trend also affect research in electronic form (5, 9). Among the reasons for lower participation of potential respondents are higher requirements to participants, decreasing selflessness of surveyed, over surveying phenomenon as well as the thorny subject of the study (particularly in relation to people diagnosed with cancer) (5, 7, 9). Difficulties in recruitment are also the result of progressive mobility perceived as a frequent change of place of residence (address), alteration or cancellation of telephone number and widespread of cellphones (5). Often traditional recruitment method fails, and the need for re-invitations to the study generates significant costs.

Given the above mentioned problems, it became necessary to search for and to implement new, efficient but also cheaper solutions. The use of the Internet as a tool to conduct questionnaire-based research started in the late 1990s of the twentieth-century. As of 2008-2009, percentage of those studies among all the methods used in epidemiological studies was at the level of 1.3% (5). The proposed solution is effective and may replace traditional ways of performing interviews (paper or telephone-based), including the assessment of the awareness of cancer prevention, as evidenced by numerous publications (4, 10, 14-17). It is noteworthy that this new strategy of researching offers the possibility of acquiring good quality data from respondents of different geographic locations. In addition, it is worth noting that both cost-effectiveness and time efficacies are better in comparison to conventional studies (5, 9, 17). These solutions allow the direct import of data to the statistical software and immediate analysis (10, 11, 18). There is the evidence for the usefulness of the Internet in prospective studies: cohort (during recruitment as well as follow-up) (5) and in research aimed at interventions (10-14).

The results of the meta-analysis presented by *Cugelman* may provide a rationalization for the effective conducting of studies involving WWW (World-Wide-Web) technologies (10). Authors have demonstrated the effectiveness of measures aimed at changing people's behaviors, including sharing of personalized

## OMÓWIENIE

**Informacje wstępne**

Jeszcze do niedawna w badaniach przekrojowych dominowała tradycyjna forma ankietowania badanych wykonywana przez wywiady (w tym wywiady telefoniczne) prowadzone przez wykwalifikowanych ankieterów. Zalety i wady ich stosowania w naukach medycznych zostały wyczerpująco omówione w znanych opracowaniach z zakresu epidemiologii (3, 5). Niestety, w ciągu ostatnich lat obserwuje się tendencję wzrostu liczby odmów potencjalnych uczestników do wzięcia udziału w badaniach ankietowych, co przekłada się na utrudnienia związane z poziomem realizacji próby (5, 8). Tendencja ta widoczna jest także w badaniach prowadzonych drogą elektroniczną (5, 9). Wśród przyczyn mniejszego zainteresowania respondentów wymienia się m.in.: wyższe wymagania stawiane badanym, coraz mniejszą bezinteresowność badanych, zjawisko przesytu badaniami kwestionariuszowymi (*oversurveying*), a także drażliwą tematykę badania szczególnie w odniesieniu do osób z rozpoznaniem choroby nowotworowej (5, 7, 9). Utrudniona rekrutacja jest także efektem postępującej mobilności, manifestującej się m.in. częstą zmianą miejsca zamieszkania (adresu), zmianą lub rezygnacją z numeru telefonu oraz rozpowszechnieniem telefonii komórkowej (5). Często tradycyjny sposób rekrutacji zawodzi, a konieczność ponownych zaproszeń do badania generuje znaczne koszty.

Wobec wspomnianych problemów konieczne stało się poszukiwanie i wdrażanie nowych, skutecznych ale i tańszych rozwiązań. Wykorzystanie sieci Internet jako narzędzia do prowadzenia badań kwestionariuszowych zapoczątkowano pod koniec lat 90. XX-tego wieku. Według danych z lat 2008-2009 ich udział wśród wszystkich metod stosowanych w badaniach epidemiologicznych kształtował się na poziomie 1,3% (5). Proponowane rozwiązanie jest skuteczne i może zastąpić tradycyjne formy zbierania wywiadów tj. papierową czy telefoniczną, czego dowodzą liczne publikacje, w tym dotyczące oceny świadomości prewencji chorób nowotworowych (4, 10, 14-17). Na uwagę zasługuje fakt, że nowy sposób prowadzenia badań oferuje możliwość pozyskania dobrej jakości danych, pochodzących od badanych z różnych lokalizacji geograficznych. Ponadto warto zwrócić uwagę na lepszą efektywność kosztową oraz czasową w porównaniu do badań konwencjonalnych (5, 9, 17). Stosowane rozwiązania umożliwiają bowiem bezpośredni import danych do programu statystycznego, w którym będą prowadzone analizy (10, 11, 18). Istnieją dowody przemawiające za przydatnością stosowania Internetu w badaniach prospektywnych: kohortowych (zarówno na etapie rekrutacji jak i obserwacji) (5) oraz w badaniach ukierunkowanych na działania interwencyjne (10-14).

Wyniki meta-analizy prezentowane przez *Cugelmana* mogą stanowić uzasadnienie dla efektywnego

materials. They pointed out the advantages of using the online technologies, including the ability to reach a wide audience at low costs. *Bottorff* and colleagues elucidated the possibilities of reaching out to young people through modern technology (including social media) with a message about health-related knowledge and ability to influence their health awareness (14).

### Models of recruitment

The recruitment of participants to studies conducted on-line can be done using either an open or closed model.

The concept of closed model involves obtaining contact details of potential respondents from sources such as census records or hospital registries (5, 20). Invitations to participate in the study are distributed using the traditional mail services, e-mail or by phone. The advantage of a closed model is a limited risk of incidental participation in the study (due to individualization of the login process), and exclusion the risk of re-participation of the same person. However, such an individualized and limited access to the platform may also discourage the respondents from taking additional actions such as logging on to the platform (5).

In the open model, information about the study is distributed to the target population using classical advertising campaigns. It is possible to use solutions to help reach the target groups for age, sex or place of residence (5, 20). A significant limitation of "open recruitment" is the inability to calculate the response rate, as well as difficulties in obtaining a representative sample; this is often associated with the possibility of selection bias. Although this issue has been discussed in the literature (21), there is a consensus that it should not be ignored (5, 9, 16). In the open model, it is difficult to eliminate the risk of multiple participation by the same person. Literature availability, as well as our own experience indicate that this phenomenon can be controlled through the use of user authentication, for example internet cookies or the registration of the IP address (Internet Protocol address) (5).

The use of social media such as Facebook is the most modern way of recruiting subjects as well as distributing information concerning cancer prevention, e.g. breast cancer (14, 22). With the possibility of direct communication to the specific target groups (targeting) their use may be useful in relation to younger population (8, 13, 14).

Consent to participate in the online survey can be expressed either in electronic as well as in traditional paper form (5, 6). Given the sensitivity of the collected personal data, it is necessary to obtain the agreement of the relevant bioethics committee or at least query this institution. Already at the design stage of the study

prowadzenia tego ostatniego rodzaju badań z udziałem technologii WWW (*World-Wide-Web* – ogólnosiwiatowa sieć) (10). W cytowanej pracy wykazano skuteczność działań ukierunkowanych na zmiany zachowań ludzi, w tym z zastosowaniem personalizacji udostępnianych materiałów. Autorzy zwrócili uwagę na zalety wykorzystania technologii *on-line*, w tym możliwość dotarcia do szerokiego grona odbiorców przy jednocześnie niskich kosztach. *Bottorff* i *wsp.* wskazali na możliwość dotarcia do osób młodych za pośrednictwem nowoczesnych technologii (w tym mediów społecznościowych) z treściami w zakresie wiedzy prozdrowotnej i możliwości kształtowania ich świadomości (14).

### Modele rekrutacji

Rekrutacja uczestników do badań realizowanych drogą *on-line* można odbywać się w modelu otwartym lub zamkniętym.

Koncepcja modelu zamkniętego zakłada pozyskiwanie danych kontaktowych potencjalnych respondentów ze źródeł takich jak spisy ludności czy rejestry szpitalne (5, 20). Zaproszenia do udziału w badaniu kierowane są za pośrednictwem poczty konwencjonalnej, elektronicznej bądź drogą telefoniczną. Zaletą modelu zamkniętego, z uwagi na indywidualne dane dostępne, jest ograniczone ryzyko uczestnictwa w badaniu osób postronnych, a także wielokrotnego udziału w badaniu tego samego respondenta. Ograniczony, zindywidualizowany dostęp do platformy może jednocześnie zniechęcać badanych do podjęcia dodatkowych działań takich jak logowanie się do platformy (5).

W modelu otwartym informacja o badaniu kierowana jest do ogółu populacji docelowej, najczęściej z wykorzystaniem klasycznych kampanii reklamowych. Możliwe jest stosowanie rozwiązań ułatwiających skierowanie przekazu do wybranych grup docelowych uwzględniających wiek, płeć, czy miejsce zamieszkania badanych (5, 20). Istotnym ograniczeniem otwartego modelu rekrutacji jest brak możliwości obliczenia wskaźnika odpowiedzi (tzw. wskaźnika partycypacji), a także problemy z uzyskaniem reprezentatywności próby z uwagi na ryzyko wystąpienia błędu selekcji. Choć kwestia ta jest dyskutowana w piśmiennictwie (21), to istnieje zgoda co do tego, iż nie należy go ignorować (5, 9, 16). W modelu otwartym trudno wyeliminować ryzyko wielokrotnej partycypacji tej samej osoby w badaniu. Dostępne dane literaturowe, a także doświadczenia własne wskazują, że zjawisko to można kontrolować poprzez stosowanie identyfikacji użytkownika np. z wykorzystaniem tzw. ciasteczek (*cookies*), a także w oparciu o rejestrację adresu IP (*Internet Protocol address*) (5).

Najnowszym sposobem rekrutacji badanych, a zarazem dystrybucji informacji dotyczących profilaktyki chorób nowotworowych, np. raka piersi jest wykorzystywanie mediów społecznościowych, takich jak *Facebook* (14, 22). Dzięki możliwości kierowania prze-

the possibility of the use of cookies or the scope and method of recording data that can be used to identify respondents by their IP address, browser type, etc. should be taken into account. An opinionated idea states that in order to protect the identity of respondents, the registration of an IP addresses could be omitted (6), while other authors suggest storing IPs in separate resources (outside the main resource in which the survey data are collected). (23).

### General principles of constructing the online questionnaires

The technical side of the questions' presented in the electronic version of the questionnaire is an important issue addressed in the literature, for example the use of the appropriate HTML (HyperText Markup Language) tags. HTML permits designing questionnaire involving close and open ended formats, while allowing at the same time, to identify the difficult questions and also to determine the optimal order of questions for improving the research (5). The use of an electronic version seems to be more appropriate in the case of studies conducted among large groups, with numbers exceeding 300 people, while for smaller numbers the traditional paper version of the questionnaire seems to be better option (20). The decision to use the Internet to conduct questionnaire-based research should be preceded by proper selection adapted to the characteristics of the study and research platform. It is possible to use ready-free of charge made solutions. An example of a free solution, which was successfully used to evaluate the awareness of Internet users in the prevention of cutaneous melanoma is Google Docs platform. This facilitated the presentation of the survey questions as well as storing the responses (6). Alternatively, original solutions can be used. However, this option is usually associated with the need to incur additional costs and time but can allow to adjust a tool for a specific research project. This problem is relatively new thereby leading to scarcity in literature in this area. Most modern solutions used to create web applications (as a research tool) offers the ability to incorporate additional innovations, along with the individualization of the questionnaire's content based on the scope of the information obtained from the examined in the preliminary stage of the investigation or solutions for validating the tool as well as the possibility of personalizing feedback (5). Electronic form of the questionnaire gives the opportunity of making the necessary corrections of errors in the content or tool's design, which is a big advantage compared to the traditional way of surveying (5). Electronic questionnaires may include additional explanatory information in the form of pop-up windows, or drop-down menus, presented directly

kazu do określonych grup docelowych (*targeting*) ich stosowanie może okazać się użyteczne w odniesieniu do osób młodych (8, 13, 14).

Deklaracja zgody na udział w badaniu internetowym może zostać wyrażona zarówno w postaci elektronicznej, jak i w tradycyjnej formie papierowej (5, 6). Z uwagi na wrażliwość zbieranych danych osobowych niezbędne jest uzyskanie zgody właściwej komisji bioetycznej albo choćby złożenie stosownego wniosku z zapytaniem. Już na etapie konstrukcji badania należy uwzględnić możliwość wykorzystania plików cookies czy zakresu i sposobu rejestrowania danych mogących posłużyć do identyfikacji respondentów poprzez ich adres IP, rodzaj przeglądarki internetowej, etc.). Istnieją opinie, że dla ochrony tożsamości badanych można pominąć rejestrację adresu internetowego IP (6), inni autorzy sugerują przechowywanie adresu w odrębnych zasobach (poza zakresem, w którym gromadzone są dane ankietowe) (23).

### Ogólne zasady budowy kwestionariuszy internetowych

Ważną kwestią poruszaną w piśmiennictwie przedmiotu jest także techniczna strona formułowania pytań stosowanych w elektronicznej wersji kwestionariusza np. z udziałem odpowiednich znaczników języka HTML (*HyperText Markup Language*). Umożliwiają one konstrukcję kwestionariusza z udziałem pytań zamkniętych, jak i otwartych, a jednocześnie pozwalają na identyfikację tzw. trudnych pytań oraz ustalenie ich optymalnej kolejności dla usprawnienia badania (5). Uznaje się, że zastosowanie wersji elektronicznej wydaje się być bardziej właściwe w przypadku badań prowadzonych w dużych grupach, o liczebności przekraczającej 300 osób, podczas gdy dla mniejszych liczebności lepszym rozwiązaniem jest tradycyjna, papierowa wersja kwestionariusza (20). Podjęcie decyzji na temat wykorzystania sieci Internet do prowadzenia badań kwestionariuszowych powinno być poprzedzone wyborem właściwej, dostosowanej do charakterystyki badania, platformy badawczej. Możliwe jest skorzystanie z gotowych rozwiązań – zarówno odpłatnych, jak i darmowych. Przykład dla rozwiązania darmowego, które z powodzeniem wykorzystano do oceny świadomości użytkowników Internetu w zakresie profilaktyki czerniaka skóry, może stanowić wykorzystanie platformy *Google Docs*, ułatwiające prezentację pytań kwestionariusza, jak i zapisywanie uzyskanych odpowiedzi (6). Alternatywnie można zastosować rozwiązania autorskie, które wiążą się zazwyczaj z koniecznością poniesienia dodatkowych nakładów finansowych i czasowych, jednakże ich stosowanie może prowadzić do optymalizacji narzędzia badawczego pod kątem specyficznego projektu badawczego. Problem jest stosunkowo nowy, a więc i dostępne piśmiennictwo w tym zakresie jest dość skąpe. Większość współczesnych rozwiązań stosowanych do tworzenia aplikacji internetowych (tu: narzędzia badawczego) oferuje

on the questionnaire page. They can also be presented as multimedia content (sound recordings, videos, animations, images) enhancing the attractiveness of the study as well as revealing the level of completeness (5, 20). Another advantage of the electronic questionnaires is almost instant access to the respondents' answers. It is worth noting that the participation in the online survey is also less time-consuming comparing with studies conducted by telephone. Additionally, subjects have the opportunity to correct their replies before submission (5). These are important arguments which justify the evaluation of cancer prevention's awareness via the Internet.

Web-based surveys offer the possibility of instant verification of completeness or proper content of the data entered by the respondent and (in case of errors) there may be an announcement of the corrections to be made (5, 20). Such functionalities are not available for the questionnaires conducted in the traditional way. Current literature indicates that the risk associated with the possibility of questions left unanswered or variant of "do not know" is lower for electronic questionnaires compared to the traditional form (5). However, the opinions of some authors indicating a greater involvement of the respondents to filling the paper-based questionnaire than electronic cannot be ignored (20).

As mentioned earlier, it is considered that the information obtained through the Internet do not differ significantly when compared to studies using traditional sources. However, the issue of reliability and representativeness of the data collected from Internet users for the general population is subject to debate (9, 24). One of the measures used to evaluate the quality of survey-based researches is the response rate which is defined as the ratio of the number of people who joined the study to the total number of invitees. The use of web technology, compared to the conventional methods, is associated with a slightly lower value for this ratio, even though these differences are not always statistically significant (5, 9, 25). Only the results of one of the published studies, regarding the aspects of the fertility of patients with a history of childhood cancer treatment, revealed that the response rate was high, primarily explained by the respondents' age homogeneity and young population (7). The authors also suggested that the use of Internet solutions is most appropriate in relation to the younger age population. This sensitive subject may contribute to a lower response rate. Overview of the available literature indicates that the response rate for the studies conducted using the Internet is extremely varied and contains the range from 0.24% to nearly 100% (20). The level of involvement of potential respondents may be associated with the use of invitations in the form

możliwość włączenia dodatkowych innowacji, wraz z indywidualizacją treści kwestionariusza w oparciu o zakres informacji uzyskanych od badanego we wstępnym etapie badania, czy też rozwiązania umożliwiające walidację narzędzia, a także możliwość spersonalizowanej informacji zwrotnej (5). Elektroniczna forma kwestionariusza daje możliwość dokonywania niezbędnych korekt błędów w zakresie treści czy konstrukcji narzędzia, co w zestawieniu z tradycyjną formą badań ankietowych jest niewątpliwą zaletą (5). Elektroniczne kwestionariusze mogą zawierać dodatkowe informacje objaśniające, w formie pojawiających się tzw. wyskakujących okien (*popup windows*), czy rozwijalnych elementów prezentowanych bezpośrednio na stronie zawierającej kwestionariusz. Mogą to być także treści multimedialne (nagrania dźwiękowe, filmowe, animacje, obrazy) podnoszące zarówno atrakcyjność projektu badawczego, czy ujawniające stopień zaawansowania udziału w badaniu (5, 20). Kolejną zaletą prowadzenia badań kwestionariuszowych w formie elektronicznej jest praktycznie niemal natychmiastowy dostęp do danych opisujących odpowiedzi respondentów. Warto podkreślić, że udział w ankiecie internetowej jest także mniej czasochłonny niż udział w badaniu prowadzonym w formie telefonicznej, dodatkowo badani mają możliwość dokonania korekty udzielonej odpowiedzi przed jej ostatecznym przesłaniem (5). To istotne argumenty przemawiające za zasadnością prowadzenia badań dotyczących oceny świadomości w zakresie profilaktyki chorób nowotworowych drogą internetową.

Badania realizowane z wykorzystaniem technologii *Web* oferują możliwość bezpośredniej weryfikacji danych wprowadzanych przez badanego pod kątem kompletności lub właściwej treści i w razie stwierdzenia nieprawidłowości informowania badanego o potrzebie dokonania korekty (5, 20). Takich możliwości nie ma w przypadku badań kwestionariuszowych prowadzonych drogą tradycyjną. Dotychczasowe dane wskazują, że zagrożenie związane z możliwością pytań pozostawionych bez odpowiedzi lub wariantu 'nie wiem' jest zdecydowanie mniejsze przy elektronicznej formie kwestionariusza niż tradycyjnej (5). Nie można jednak pominąć opinii innych autorów wskazujących na większe zaangażowanie badanych przy wypełnieniu kwestionariusza w formie papierowej niż jego elektronicznego odpowiednika (20).

Jak już wspomniano, uznaje się, że informacje pozyskiwane drogą internetową nie odbiegają istotnie od tych pochodzących z badań wykorzystujących rozwiązania tradycyjne. Jednak kwestia rzetelności, wiarygodności oraz reprezentatywności danych pozyskiwanych od użytkowników sieci Internet dla populacji generalnej poddawana jest dyskusji (9, 24). Jedną z miar wykorzystywanych do oceny jakości badania kwestionariuszowego jest odsetek odpowiedzi (*response rate*) czyli stosunek liczby osób, które przystąpiły do badania do łącznej liczby zaproszonych. W porównaniu do metod konwencjonalnych,

of personal letters addressed via e-mail (18, 20). The authority of the research team, as well as the interest of the respondent convinced, allows the topic of the study to be significant (7, 26). Sophisticated tools' design and time-consuming filling of the survey could impact negatively on the participation rate. To ensure an appropriate level of accuracy and attractiveness of materials shared via a network, cooperation is essential between the research team of programmers and with the target group. The substantive issues also require to carry out pilot studies (27), which is particularly recommended for researches on medical topics, including these cancer-related. It is necessary to adapt the survey to the level of participants' Internet skills at the design stage (23).

### **Characteristics of the on-line studies' participants and evaluation of representativeness**

Internet users and participants of on-line epidemiological studies, are often young people (4, 5, 16). It could be considered to be an important accusation in relation to the method of respondents' selection, as this method is prone to a specific type of bias called a selection bias. In the majority of cited papers, with the exception of *Fleischer* (16) and *Shahrokni* (19), significant differences between the population recruited via the Internet and the general population were not confirmed in terms of age and socio-economic profile (5). Patients after cancer treatment who were examined on-line differed from those surveyed traditionally in the field of quality of life (better), income (higher), age (younger), educational level (better), and place of residence (more urban areas) (19). It was also found that Internet users significantly more often trust and paid more attention to Internet sources compared to those not benefiting from access to the global network. It is worth noting that both groups of respondents indicated that the physician is their most reliable source of information about health (19).

An important reason for the lack of representativeness of the respondents in relation to the general population is selection bias (3). Recent studies have indicated that the importance of this bias in the web-based studies systematically decreases (2). It was also observed that selection bias is a high burden in relation to observation of quantitative than qualitative studies (21). Because this type of error can lead to the complete omitting of certain phenomena, and consequently to the incorrect description of the situation, it should be reckon with the consequences for the final inference (5, 9, 16). Moreover, among the shortcomings of surveys conducted electronically, the error resulting from intent to enter data directly by the subjects themselves (reported bias) (28) should be included. Answering questions in line with social

zastosowanie technologii internetowej wiąże się z uzyskaniem nieco niższych wartości wskaźnika, choć nie zawsze różnice te są istotne statystycznie (5, 9, 25). Tylko wyniki jednego z opublikowanych badań, które dotyczyło aspektów płodności po przebytych w dzieciństwie leczeniu onkologicznym, wskazały, że odsetek odpowiedzi był wysoki, co sami autorzy tłumaczą jednorodnością badanych pod względem wieku i raczej młodą populacją (7). Autorzy sugerują także, że w odniesieniu do młodszych wiekowo populacji stosowanie rozwiązań internetowych jest jak najbardziej właściwe. Wskazują jednocześnie, że z uwagi na drażliwą tematykę badania wskaźnik uzyskanych odpowiedzi będzie niski. Przegląd dostępnego piśmiennictwa wskazuje, że wartości współczynnika odpowiedzi dla badań prowadzonych z udziałem Internetu są niezwykle zróżnicowane i kształtują się w zakresie od 0,24% do blisko 100% (20). Na zaangażowanie się potencjalnych respondentów wpływać może stosowanie zaproszeń w formie indywidualnych listów kierowanych do badanych za pośrednictwem poczty elektronicznej (18, 20). Nie bez znaczenia jest także autorytet zespołu badawczego oraz przekonanie respondentów o przydatności badania, czy po prostu zainteresowanie badanych poruszaną tematyką (7, 26). Skomplikowana konstrukcja narzędzia i czasochłonność jego wypełniania wpływają negatywnie na odsetek partycypacji. Dla zapewnienia odpowiedniego poziomu poprawności i atrakcyjności udostępnianych za pośrednictwem sieci materiałów niezbędna jest współpraca zespołu badawczego z programistami oraz z grupą docelową. Kwestie merytoryczne wymagają również przeprowadzenia badań pilotażowych (27), co ma szczególne uzasadnienie w przypadku badań dotyczących tematyki medycznej, w tym chorób nowotworowych. Na etapie projektowania niezbędne jest dostosowanie badania do poziomu umiejętności uczestników badania w zakresie użytkowania sieci (23).

### **Charakterystyka uczestników badań on-line i ocena reprezentatywności**

Użytkownikami sieci Internet, a w związku z tym także uczestnikami badań epidemiologicznych realizowanych drogą *on-line* częściej są osoby młode (4, 5, 16). Można uznać, że jest to istotny zarzut w odniesieniu do stosowanej metody doboru badanych, metoda obarczona jest swoistym rodzajem błędu zwanym błędem selekcji (*selection bias*). W większości cytowanych prac, z wyjątkiem pracy *Fleischer'a* (16) i *Shahrokni* (19), nie potwierdzono istotnych różnic w zakresie wieku i profilu społeczno-ekonomicznego pomiędzy populacją badaną pozyskaną z wykorzystaniem Internetu a populacją generalną (5). Zbadani drogą *on-line* pacjenci po zakończeniu leczenia przeciwnowotworowego różnili się od badanych metodami tradycyjnymi w zakresie jakości życia (lepsz), dochodów (wyższe), wieku (młodszy), poziomu wykształcenia (lepsz), a także miejscem zamieszkania

expectations (social desirability bias) is less common in the case of electronic questionnaires, which is important especially in the case of research on so-called sensitive issues as sexuality, the use of psychoactive substances, etc. The authors explain that it may have something to do with a sense of greater anonymity, which leads to a more sincere, and not socially expected, responses (5). The completeness of the questionnaire, assessed as a completion rate, affects the quality of the data obtained, which is also important in the case of research projects using new technologies. The main reasons for incomplete answers in the questionnaire include construction errors, extensive questionnaire (associated with the need to scroll the window down, of which respondents may not be aware); the lack of compatibility with the device used to display the content; too difficult questions; superficial and / or too hasty reading of the questions (5). Extensive, multi-page questionnaires usually require long-completion time and deteriorates response rate. In contrast it may lead to a better quality of data. The low response rate can result in significant differences between the study group and non-responders. Such a situation may account for the potential lack of representativeness of the study group compared to the general population (non-response bias) (6). Conducting a study while maintaining privacy data protection and ethics make it impossible to assess the differences between the study participants and non-responders. To substantially improve response rates in web-based studies, all of the methods known from conventional surveys may be applied. Additionally, it may be possible to emphasize the benefits of sharing the questionnaire in PDF format as well as the user-friendly and appropriately designed website (5). Apart from the form, the length of the questionnaire, appropriate and understandable language to communicate questions are essential for greater response rate (as already mentioned in the description of the research tool construction).

Another important issue for the representativeness is the higher commitment to the study of Internet people who have been diagnosed than the healthy people. In the Anglo-Saxon literature the term cancer survivors have been used to define those ever diagnosed with or those who completed the process of cancer treatment. It is believed that the search for information is a patient's form of coping with the situation of the cancer diagnosis. Usually, these people are interested in information about prognosis, available treatment options and potential side effects of treatment and the impact of disease and undertaken treatment on quality of life (17-19, 29).

(częściej obszary zurbanizowane) (19). Stwierdzono również, że badani internauci istotnie częściej ufali źródłom internetowym i poświęcali im więcej uwagi w porównaniu do osób niekorzystających z dostępu do sieci globalnej. Warto podkreślić, iż obydwie grupy badanych wskazywały, że lekarz jest dla nich najbardziej wiarygodnym źródłem informacji o zdrowiu (19).

Istotną przyczyną braku reprezentatywności badanych w stosunku do populacji generalnej jest wspomniany już błąd selekcji (3). Najnowsze prace wskazują, że systematycznie zmniejsza się znaczenie omawianego błędu w badaniach prowadzonych w oparciu o dostęp do sieci Internet (2). Zaobserwowano także, że błąd selekcji jest większym obciążeniem w odniesieniu do obserwacji ilościowych, niż badań jakościowych (*qualitative studies*) (21). Ponieważ błąd ten może prowadzić do całkowitego pominięcia pewnych zjawisk, a w konsekwencji do nieprawidłowego opisu sytuacji, to należy w dalszym ciągu liczyć się z jego konsekwencjami dla ostatecznego wnioskania (5, 9, 16). Ponadto, wśród mankamentów badań ankietowych prowadzonych drogą elektroniczną wymienić należy błąd wynikający z intencyjnego podawania danych bezpośrednio przez samych badanych (*reported bias*) (28). Natomiast udzielanie odpowiedzi zgodnych ze społecznymi oczekiwaniami (*social desirability bias*) jest rzadziej spotykane w przypadku kwestionariuszy elektronicznych, co ma znaczenie zwłaszcza w przypadku badań poruszających tzw. wrażliwą tematykę dotyczącą seksualności, stosowania środków psychoaktywnych, etc. Autorzy tłumaczą, że prawdopodobnie ma to związek z poczuciem większej anonimowości, co prowadzi do udzielania bardziej szczerych, a nie społecznie poprawnych odpowiedzi (5). Kompletność wypełnienia kwestionariusza badawczego, którego miarą jest wskaźnik kompletności (*completion rate*) wpływa na jakość pozyskiwanych danych, co ma miejsce także w przypadku projektów badawczych wykorzystujących nowoczesne technologie. Wśród głównych przyczyn niekompletnych odpowiedzi w badaniach kwestionariuszowych znajdują się m.in. błędy konstrukcyjne związane ze zbyt obszernym kwestionariuszem (konieczność przewinięcia okna w dół, czego badany nie musi być świadomy); z brakiem kompatybilności urządzenia wykorzystywanego do wyświetlania treści kwestionariusza przez badanego; z zastosowaniem zbyt trudnych pytań; z pobieżną i/lub zbyt pośpieszną lekturą treści pytań (5). Obszerny, wielostronicowy kwestionariusz wiąże się z wydłużeniem czasu niezbędnego do jego wypełnienia, a w konsekwencji pogarsza współczynnik odpowiedzi. Z drugiej strony może jednak prowadzić do uzyskania danych o lepszej jakości. Niski odsetek odpowiedzi może powodować istotne różnice pomiędzy grupą badaną a osobami, które nie zdecydowały się na wzięcie udziału w badaniu. Taka sytuacja może odpowiadać za potencjalny brak reprezentatywności badanej grupy w stosunku do populacji gene-



### Data security

Data security is an important aspect of on-line surveys. The risk of third party access to data (often sensitive) stored during the survey should be taken into an account at the design stage of the study (choice of platform). It appears to be particularly important for collecting data in the field of medical history, former therapies, including cancer treatment (7). It seems to be important for ethical as well as legal issues. It should be determined whether the risk associated with the use of an external (often paid) research software does not exceed the expected benefits (12, 23). On the basis of the literature and authors' own experience, it should be stated that in order to limit the risk of third party access to the acquainted data, it is recommended to use encrypted Internet connections (HTTPS - Hypertext Transfer Protocol Secure), appropriately secured databases as well as the firewall protection (5, 23).

### CONCLUSIONS

Literature which are available allowed to conclude that the web-based questionnaire studies could be an effective solution and may replace traditionally conducted interviews. It appears that an on-line mode is more adequate for querying young people. For the development of modern research techniques it is also important to promote active network access among the elderly by improving their ability to use online resources. It is necessary to continue efforts to standardize the studies conducted on-line with an assessment of their strengths and weaknesses.

### REFERENCES

1. Didkowska J, Wojciechowska U. *Zachorowania i zgony na nowotwory złośliwe w Polsce*. Krajowy Rejestr Nowotworów, Centrum Onkologii - Instytut im. Marii Skłodowskiej - Curie. <http://onkologia.org.pl/k/epidemiologia/> Accessed: 27.12.2015
2. Główny Urząd Statystyczny. *Spółeczeństwo informacyjne w Polsce*. [http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5497/2/5/1/spoleczenstwo\\_informacyjne\\_w\\_polsce\\_2015\\_-\\_notatka.pdf](http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5497/2/5/1/spoleczenstwo_informacyjne_w_polsce_2015_-_notatka.pdf) Accessed: 27.12.2015
3. Jędrzychowski W. *Epidemiologia kliniczna w medycynie i zdrowiu publicznym*. Podręcznik dla studentów i lekarzy. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego; 2010:180-183.
4. Miyagi E, Motoki Y, Asai-Sato M, et al. Web-based recruiting for a survey on knowledge and awareness of cervical cancer prevention among young women living in Kanagawa prefecture, Japan. *Int J Gynecol Cancer*. 2014;24:1347-55.

ralnej (*non-response bias*) (6). Prawidłowo prowadzone badanie z zachowaniem polityki prywatności, ochrony danych osobowych i względów etycznych praktycznie uniemożliwia jednoznaczną ocenę różnic pomiędzy uczestnikami badania a tymi osobami, które odmówiły udziału w badaniu (*non-responders*). Dla poprawy odsetka odpowiedzi w badaniach internetowych zastosowanie mogą mieć praktycznie wszystkie metody znane z badań realizowanych metodami tradycyjnymi. Dodatkowo podkreśla się korzyści wynikające z udostępnienia kwestionariusza w formie pliku PDF oraz właściwej, przyjaznej dla użytkownika formy zaprojektowanej strony internetowej (5). Poza formą, istotna dla większego odsetka odpowiedzi jest także długość samego kwestionariusza badawczego, właściwy i zrozumiały język komunikowania pytań, o czym już wspomniano przy opisie konstrukcji narzędzia badawczego.

Kolejna ważna dla reprezentatywności kwestia to większe zaangażowanie w badanie internetowe osób z rozpoznaną chorobą, niż osób zdrowych. W piśmiennictwie anglosaskim używa się określenia *cancer survivors* dla określenia osób z postawionym kiedykolwiek rozpoznaniem choroby nowotworowej lub pacjentów, u których zakończono proces leczenia choroby nowotworowej. Uważa się, że poszukiwanie informacji jest pewną formą radzenia sobie chorego z sytuacją rozpoznania nowotworu. Najczęściej w kręgu zainteresowań tych osób znajdują się zagadnienia takie jak: informacje o rokowaniu, o dostępnych opcjach terapeutycznych i potencjalnych skutkach ubocznych leczenia oraz wpływie choroby i podjętego leczenia na jakość życia (17-19, 29).

### Bezpieczeństwo danych

Istotnym aspektem prowadzenia badań kwestionariuszowych w modelu *on-line* jest bezpieczeństwo danych. Na etapie projektowania badania (wybór platformy) należy uwzględnić ryzyko dostępu osób trzecich do często poufnych danych rejestrowanych zawartych w ankiecie, co wydaje się szczególnie istotne w przypadku gromadzenia danych z zakresu wywiadu chorobowego i przebytego leczenia, w tym np. przeciwnowotworowego (7) Jest to istotne ze względów etycznych, ale i prawnych. Należy określić, czy ryzyko związane z wykorzystaniem zewnętrznego (często płatnego) oprogramowania badawczego nie przewyższa spodziewanych korzyści (12, 23). Na podstawie piśmiennictwa oraz doświadczeń własnych autorów należy stwierdzić, iż celem ograniczenia dostępu osób trzecich do rejestrowanych danych zalecane jest wykorzystanie szyfrowanych połączeń internetowych (protokół HTTPS – *Hypertext Transfer Protocol Secure*) oraz odpowiednie zabezpieczenia bazy danych, w tym stosowanie tzw. zapory ogniowej (*firewall*) (5, 23).

## WNIOSKI

5. van Gelder MMHJ, Bretveld RW, Roeleveld N. Web-based Questionnaires: The Future in Epidemiology? *Am J Epidemiol.* 2010;172:1292-8.
6. Kamińska-Winciorek G, Gajda M, Wydmański J, et al. What do Web users know about skin self-examination and melanoma symptoms? *Asian Pac J Cancer Prev.* 2015;16:3051-6.
7. van den Berg MH, Overbeek A, van der Pal HJ, et al. Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics. *J Med Internet Res.* 2011;13:e76.
8. Fenner Y, Garland SM, Moore EE, et al. Web-Based Recruiting for Health Research Using a Social Networking Site: An Exploratory Study. *J Med Internet Res.* 2012;14:e20.
9. Hohwü L, Lyshol H, Gissler M, et al. Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. *J Med Internet Res.* 2013;15:e173.
10. Cugelman B, Thelwall M, Dawes P. Online Interventions for Social Marketing Health Behavior Change Campaigns: A Meta-Analysis of Psychological Architectures and Adherence Factors. *J Med Internet Res.* 2011;13:e17.
11. Kaplowitz MD, Hadlock TD, Levine R. A Comparison of Web and Mail Survey Response Rates. *Public Opin Q.* 2004;68:94-101.
12. Ahern NR. Using the Internet to conduct research. *Nurse Res.* 2005;13:55-70.
13. Richardson CG, Struik LL, Johnson KC, et al. Initial Impact of Tailored Web-Based Messages About Cigarette Smoke and Breast Cancer Risk on Boys' and Girls' Risk Perceptions and Information Seeking: Randomized Controlled Trial. *JMIR Res Protoc.* 2013;2:e53.
14. Bottorff JL, Struik LL, Bissell LJJ, et al. A social media approach to inform youth about breast cancer and smoking: An exploratory descriptive study. *Collegian.* 2014;21:159-68.
15. Webb TL, Joseph J, Yardley L, et al. Using the Internet to Promote Health Behavior Change: A Systematic Review and Meta-analysis of the Impact of Theoretical Basis, Use of Behavior Change Techniques, and Mode of Delivery on Efficacy. *J Med Internet Res.* 2010;12:e4.
16. Fleisher L, Kandadai V, Keenan E, et al. Build It, and Will They Come? Unexpected Findings From a Study on a Web-Based Intervention to Improve Colorectal Cancer Screening. *J Health Commun.* 2012;17:41-53.
17. Goode AD, Lawler SP, Brakenridge CL, et al. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. *J Cancer Surviv.* 2015;9:660-82.
18. Short CE, Rebar AL, Vandelanotte C. Do personalised e-mail invitations increase the response rates of breast cancer survivors invited to participate in a web-based behaviour change intervention? A quasi-randomised 2-arm controlled trial. *BMC Med Res Methodol.* 2015;15:66.
19. Shahrokni A, Mahmoudzadeh S, Lu BT. In Whom Do Cancer Survivors Trust Online and Offline? *Asian Pacific J Cancer Prev.* 2014;15:6171-6.
20. Uhlig CE, Seitz B, Eter N, et al. Efficiencies of Internet-Based Digital and Paper-Based Scientific Surveys and the Estimated Costs and Time for Different-Sized Cohorts. van den Besselaar P, ed. *PLoS One.* 2014;9(10):e108441.
21. Greenhalgh T, Taylor R. How to read a paper: Papers that go beyond numbers (qualitative research). *BMJ.* 1997;315:740-3.
22. Springvloed L, Lechner L, de Vries H, et al. Short- and Medium-Term Efficacy of a Web-Based Computer-Tailored Nutrition Education Intervention for Adults Including Cognitive and Environmental Feedback: Randomized Controlled Trial. *J Med Internet Res.* 2015;17:e23.
23. Gill FJ, Leslie GD, Grech C, et al. Using a web-based survey tool to undertake a Delphi study: Application for nurse education research. *Nurse Educ Today.* 2013;33:1322-8.
24. Duffy ME. Methodological issues in Web-based research. *J Nurs Scholarsh.* 2002;34:83-8.
25. Zuidgeest M, Hendriks M, Koopman L, et al. A Comparison of a Postal Survey and Mixed-Mode Survey Using a Questionnaire on Patients' Experiences With Breast Care. *J Med Internet Res.* 2011;13:e68.
26. Salkind, Neil J *Encyclopedia of research design.* SAGE Publications, Thousand Oaks, Calif, 2010.
27. Jander A, Crutzen R, Mercken L, et al. A Web-based computer-tailored game to reduce binge drinking among 16 to 18 year old Dutch adolescents: development and study protocol. *BMC Public Health.* 2014;14:1054.

28. Choma K, McKeever AE. Cervical Cancer Screening in Adolescents: An Evidence-Based Internet Education Program for Practice Improvement Among Advanced Practice Nurses. *Worldviews Evidence-Based Nurs.* 2015;12:51-60.
29. Mayer DK, Terrin NC, Kreps GL, et al. Cancer survivors information seeking behaviors: A comparison of survivors who do and do not seek information about cancer. *Patient Educ Couns.* 2007;65:342-50.

Received: 26.01.2016

Accepted for publication: 25.04.2016

Otrzymano: 26.01.2016 r.

Zaakceptowano do publikacji: 25.04.2016 r.

**Adres do korespondencji:**

**Address for correspondence:**

Dr hab. n. med. Małgorzata Kowalska

Śląski Uniwersytet Medyczny w Katowicach, Katedra i Zakład

Epidemiologii WL w Katowicach, ul. Medyków 18,

40-752 Katowice-Ligota

tel. +48 32 20 88 536; tel/fax +48 32 25 23 734; e-mail:

mkowalska@sum.edu.pl