

Problem wielokrotnego testowania w przesiewowych badaniach poligraficznych – implikacje dla bezpieczeństwa wewnętrznego

Abstrakt

Badania poligraficzne o charakterze przesiewowym najczęściej przeprowadza się przed zatrudnieniem kandydata. Statystyka Bayesowska pomaga lepiej zrozumieć rzeczywiste znaczenie uzyskiwanych wyników testów. Zdarza się również wykonywanie badań powtórnych – najczęściej wtedy, gdy za pierwszym podejściem kandydatowi się nie powiedzie. W artykule autor rozważa różne scenariusze związane z ponownym testowaniem oraz ich konsekwencje, także w zakresie bezpieczeństwa wewnętrznego poszczególnych instytucji i państw. Poszukuje optymalnych rozwiązań, aby procedury testowania sprzyjały szczelności systemów, a zarazem pozostały wydajne i użyteczne.

Słowa kluczowe poligraf, badania, przesiew, powtórne testowanie, wyniki, interpretacja, bezpieczeństwo.

Badania przesiewowe w różnych obszarach życia przeprowadza się wobec osób, u których nie ujawniły się żadne konkretne oznaki nieprawidłowości, sprecyzowane problemy czy wątpliwości do weryfikacji. Badania takie umożliwiają określenie ryzyka zaistnienia lub wykrycie określonej choroby, zaburzenia bądź niepożądanego zachowania. Polegają na wstępnym rozpoznaniu problemu przed pojawieniem

się jego widocznych symptomów. Stanowią element profilaktyki, wczesnego wykrywania i prognozowania zagrożeń.

Istota i znaczenie przesiewowych badań poligraficznych w procedurach kadrowych

Przesiewowy charakter mają między innymi badania psychofizjologiczne z wykorzystaniem poligrafu przeprowadzane w ramach postępowań rekrutacyjnych do pracy lub służby, przy sprawdzaniach kontrwywiadowczych i antysabotażowych w grupach podwyższonego ryzyka (np. wśród agentów, informatorów, osób przewidzianych do specjalnych zadań). Chociaż pierwotnie metody badań poligraficznych zaprojektowano do celów śledczych, we współczesnym świecie (m.in. w USA, Chinach, Izraelu, Federacji Rosyjskiej, wielu krajach europejskich, w tym w Polsce) najczęściej badań jest realizowanych w ramach procedur kadrowych – zarówno w instytucjach państwowych, jak i w sektorze prywatnym. Przede wszystkim wykonuje się je w rządowych agencjach bezpieczeństwa, organach egzekwowania prawa i innych podmiotach, których działalność ma istotne znaczenie dla bezpieczeństwa i porządku publicznego.

Przesiewowe badania poligraficzne przeprowadzane w celach kadrowych spełniają równocześnie kilka funkcji: a) selekcyjną – począwszy od autoselekcji, tj. odstraszenia nieodpowiednich kandydatów, którzy znają lub przewidują, jakie mogą być kryteria naboru, i mają świadomość, że ich nie spełniają; b) prewencyjną – przez zapobieganie niewłaściwym aktywnościom i zniechęcanie do nich – dotyczy to zarówno tych, którzy starają się o angaż, jak i osób już zatrudnionych; c) informacyjną, wykrywczą – poszerzenie informacji o osobie badanej (w tym o takie, które nie są dostępne innymi metodami), naprowadzenie na potencjalne problemy do dalszej weryfikacji.

W literaturze wskazuje się, że wśród najbardziej odpowiednich tematów, o które kandydaci, np. na funkcjonariuszy służb mundurowych, są pytani podczas badań poligraficznych, znajdują się: kwestie związane z nietolerancją (używanie przemocy domowej, obelg na tle etnicznym i rasowym), zaangażowanie w działalność przestępczą w życiu dorosłym, zażywanie narkotyków w ostatnich kilku latach, postępowania dyscyplinarne w poprzednich miejscach pracy¹. Dobór zagadnień zależy od specyfiki danej instytucji i jej polityki wewnętrznej. Może uwzględniać między innymi ukrywanie kontaktów z grupami podwyższonego ryzyka, materiały

¹ M. Handler i in., *Integration of Pre-Employment Polygraph Screening into the Police Selection Process*, „Journal of Police and Criminal Psychology” 2009, t. 24, nr 2, s. 69–86.

ewentualnego nacisku czy sposób postępowania z informacjami niejawnymi (jeżeli wcześniej był przyznany dostęp do takich informacji).

W Polsce badania poligraficzne stanowią jeden z etapów postępowań kwalifikacyjnych do wszystkich służb specjalnych (Agencja Bezpieczeństwa Wewnętrznego, Agencja Wywiadu, Służba Kontrwywiadu Wojskowego, Służba Wywiadu Wojskowego, Centralne Biuro Antykorupcyjne), Służby Ochrony Państwa, Straży Granicznej RP, a także w wybranych komórkach Krajowej Administracji Skarbowej, w których są wykonywane czynności określone w ustawie o KAS. Jeżeli chodzi o akty normatywne na poziomie ustawy lub rozporządzenia, to jedynie w przypadku SOP, SG, KAS i ABW mamy do czynienia ze wskazaniem obligatoryjności tego etapu (zob. art. 72 ust. 1 pkt 2 *Ustawy z dnia 8 grudnia 2017 r. o Służbie Ochrony Państwa*; art. 31 ust. 1a pkt 5 *Ustawy z dnia 12 października 1990 r. o Straży Granicznej*; art. 153 ust. 2 *Ustawy z dnia 16 listopada 2016 r. o Krajowej Administracji Skarbowej*; § 5 ust. 1 pkt 6 lit. d *Rozporządzenia Prezesa Rady Ministrów z dnia 29 listopada 2002 r. w sprawie wzoru kwestionariusza osobowego oraz szczegółowych zasad i trybu przeprowadzania postępowania kwalifikacyjnego wobec kandydatów do służby w Agencji Bezpieczeństwa Wewnętrznego*). W Żandarmerii Wojskowej badaniom poligraficznym poddaje się osoby rozpatrywane na „stanowiska wymagające szczególnych predyspozycji” (zob. art. 8a ust. 8 pkt 1 lit. d *Ustawy z dnia 24 sierpnia 2001 r. o Żandarmerii Wojskowej i wojskowych organach porządkowych*). Podobne zapisy, teoretycznie ograniczające liczbę kandydatów podlegających takim badaniom, znajdują się w aktach prawnych regulujących funkcjonowanie wymienionych służb specjalnych. W rzeczywistości jednak wykaz stanowisk wymagających szczególnych predyspozycji lub umiejętności obejmuje wszystkie etaty funkcjonariuszy, czyli tzw. etaty mundurowe (z reguły nie dotyczy pracowników cywilnych). Przepisy zawierają przy tym pewne „furtki” pozwalające pominąć badanie poligraficzne, m.in. w odniesieniu do funkcjonariuszy przenoszących się pomiędzy służbami (tak jest np. w SOP – zob. art. 72 ust. 1 pkt 1 lit. b ww. ustawy o SOP; zob. także art. 50 ust. 3 *Ustawy z dnia 9 czerwca 2006 r. o Centralnym Biurze Antykorupcyjnym*).

Fakultatywnie wykonanie badań poligraficznych przed zatrudnieniem dopuszcza się również w Policji. Dotychczasowe doświadczenia nie wskazują jednak, aby rutynowo z tego instrumentu korzystano, poza wewnętrznym naborem do Biura Spraw Wewnętrznych.

Każda z wymienionych instytucji ma również podstawy prawne do poddawania badaniom poligraficznym nie tylko kandydatów, lecz także funkcjonariuszy. Mimo że z istoty tych badań wynika ich dobrowolność (a przynajmniej nieprzymusowy charakter), to w praktyce nieuzasadniona odmowa poddania się testom z wykorzystaniem poligrafu będzie potraktowana jako delikt dyscyplinarny, gdyż skierowanie na badanie (zwykle pozostające w kompetencjach szefa danej służby)

jest formą polecenia służbowego. W przypadku kandydatów badanie jest dobrowolne w tym sensie, że sama aplikacja do służby wynika z wolnej woli. Natomiast postępowanie kwalifikacyjne aplikanta odmawiającego udziału w badaniu zakończy się decyzją odmowną w kwestii przyjęcia do służby.

Osobnym zagadnieniem jest potrzeba współpracy między ekspertem przeprowadzającym badanie poligraficzne a osobą badaną. Uzyskanie miarodajnych wyników testów wymaga podążania za instrukcjami badającego. Ujawnienie symptomów celowego braku współpracy ze strony badanego (w tym przez stosowanie tzw. środków zakłócających) prowadzi do sformułowania opinii równoznacznej w skutkach z wynikami testów wskazujących na wprowadzanie w błąd. W tych okolicznościach zwykle przestaje się na jednym badaniu, bez możliwości powtórnego testowania. Ponadto jeśli odrzucenie kandydata do służby jest konsekwencją wyniku badania poligraficznego, to dzieje się tak albo z powodu dyskwalifikujących oświadczeń, które przekazał on w toku badania, albo z powodu niedających się usunąć wątpliwości co do wiarygodności odpowiedzi udzielonych w czasie testów zakończonych niekorzystnymi wynikami.

Badanie poligraficzne jako etap postępowania kwalifikacyjnego do służby mundurowej ma pomóc w ocenie predyspozycji kandydata, ale przede wszystkim dostarcza rekruterom i oficerom pionu ochrony cennego narzędzia w zarządzaniu ryzykiem w instytucji. Nie jest to narzędzie jedyne i wystarczające do podejmowania określonych decyzji kadrowych, wykorzystuje się je równolegle z innymi procedurami. Niemniej jednak, ze względu na jego efektywność w pozyskiwaniu od kandydata nieznanych dotąd informacji, a także wysoką wartość diagnostyczną samych testów, jest to etap niezwykle istotny, a ewentualne wątpliwości, które po nim pozostają, trudno tak po prostu pominąć.

W żadnej dziedzinie życia nie istnieje chyba test, który ma doskonałą czułość i swoistość, a tym samym daje sto procent dokładnych wskazań i jest perfekcyjnie użyteczny (nigdy nie przynosi wyników nierozstrzygniętych). Swoje ograniczenia mają także testy wykorzystywane w badaniach poligraficznych. Chodzi jednak o to, aby zaakceptować poziom prawdopodobieństwa, który zminimalizuje odsetek fałszywych identyfikacji, zarówno obciążających osoby niewinne, szczerze (błąd typu pierwszego), jak i odciążających osoby winne, kłamiące (błąd typu drugiego)².

Osoby kształtujące politykę danej instytucji raczej nie mają świadomości (albo rzadko ją mają), że mogą wypracować z poligraferami zasady, które określą priorytety w testowaniu, granice tolerancji błędu, czyli to, jak grube oczka może mieć to sito, aby z jednej strony było ono skuteczne, a z drugiej – aby cały proces

² Por. R. Eggleston, *Sixth Wilfred Fullagar Memorial Lecture: Beyond reasonable doubt*, „Monash University Law Review” 1977, t. 4, nr 1, s. 1–2.

przebiegał sprawnie. To mogłyby być zasady odnoszące się do wszystkich testów albo wyodrębnione dla konkretnych obszarów tematycznych poruszanych w badaniach. Bardziej restrykcyjną politykę należałoby rozważyć na przykład w odniesieniu do zagadnień kontrwywiadowczych. Poligraferzy dalej poszukiwaliby tych samych cech diagnostycznych zapisów na poligramach, jedyna zmiana dotyczyłaby reguł decyzyjnych (progów ocen numerycznych), przekładających się na końcowe wyniki testów. To jest problem wymagający odpowiedzi na dwa idące w parze pytania: 1) jaki odsetek osób nieszczerych (i w jakich tematach) można zaakceptować, aby niepostrzeżenie przeszły przesiew przy bardziej liberalnych kryteriach, a w zamian nie pozbawić większej grupy z tych szczerych szansy, na którą zasługują, oraz zwiększyć liczbę osób, które przejdą do dalszego etapu rekrutacji?; 2) jak bardzo można uszczelnić procedurę, aby skuteczniej odrzucać niepożądane, nieszczerze osoby i w konsekwencji nietrafnie oceniać proporcjonalnie większą liczbę kandydatów bez zarzutu, u których pechowo wystąpiłby fałszywy alarm, oraz nieco przyblokować drożność całej procedury? Statystyka wskazuje złoty środek w tym zakresie, ale trudno stwierdzić, czy odpowiada on rzeczywistym interesom instytucji i szerzej – interesowi państwa.

Jeśli weźmie się pod uwagę wszystkie potwierdzone naukowo testy – zarówno wykonywane w celach przesiewowych, jak i diagnostycznych (jednowątkowych), to ich średnia dokładność została oszacowana na 87,1% (badania przesiewowe są dokładne na poziomie 85%, a jednowątkowe – 92,1%, przy średnich odsetkach wyników nierozstrzygniętych wynoszących odpowiednio: 12,5% i 8,8%). Metaanaliza opublikowana przez American Polygraph Association w 2011 r. dostarcza szczegółowych danych o parametrach poszczególnych testów spełniających kryteria naukowej walidacji³. Od tamtego czasu nie wykonano podobnych analiz, a wyniki nielicznych nowych badań empirycznych zostały opublikowane bądź odrębnie, bądź jako suplement do wspomnianego raportu z 2011 r.⁴

Z kolei statystyczne dane referencyjne zestawione w ramach dostępnych materiałów na temat uznanych systemów analizy danych testowych pozwalają na matematyczne przedstawienie znaczenia konkretnych rezultatów testów – określenie, z jak mocnym wynikiem mamy do czynienia, i wyrażenie stopnia pewności badającego⁵. W ten sposób wyznacza się wartości p , tj. prawdopodobieństwa uzyskania

³ American Polygraph Association, *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, „Polygraph” 2011, t. 40, nr 4.

⁴ Zob. np. *Addendum to the 2011 Meta-analytic Survey – the Utah Four-Question Test (“Raskin Technique”)/ ESS*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2020, t. 49, nr 2, s. 73–81.

⁵ Por. R. Nelson i in., *Using the Empirical Scoring System*, „Polygraph” 2011, t. 40, nr 2, s. 67–78; R. Nelson, *Multinomial Reference Distributions for the Empirical Scoring System*, „Polygraph & Forensic

określonego wyniku lub jeszcze bardziej ekstremalnego od hipotezy zerowej, która zakładałaby, że dla tego wyniku nie ma różnicy między osobami odpowiadającymi na pytania krytyczne szczerze i nieszczerze. Alternatywnie można się posłużyć kategorią szansy, która opisuje, ile razy bardziej prawdopodobna jest prawdziwość danej hipotezy niż jej fałszywość. Ale uwaga – nawet jeśli takie dane są dostępne i na przykład wskazują na wprowadzanie w błąd przy pytaniach o popełnienie jakiegoś czynu, to nie są one równoznaczne z faktycznym prawdopodobieństwem zaistnienia takiego zdarzenia. Dowiadujemy się tylko i aż, jak często uzyskany rezultat testu występuje w określonej populacji (np. osób wprowadzających w błąd). Ewentualnie tego, jak zmieniają się szanse – stosunek prawdopodobieństwa danej hipotezy do hipotezy przeciwnej – po uzyskaniu danego wyniku testu.

Nie sposób przy tym pominąć tzw. wskaźnika bazowego, mówiącego o tym, jakie wyjściowo szanse można przypisać wystąpieniu danego zdarzenia czy cechy w badanej populacji. Inne będzie prawdopodobieństwo trafnego zdiagnozowania przez psychologa jakiegoś zaburzenia osobowości podczas badań wykonywanych w grupie kandydatów do służby, a inne w populacji pacjentów szpitala psychiatrycznego. Więcej zakażonych wirusem należy się spodziewać w grupie pacjentów objawowych niż wśród bezobjawowych. Szansa na pojawienie się szpiega w populacji wędkarzy jest przypuszczalnie mniejsza niż w populacji osób starających się o zatrudnienie w miejscach związanych z dostępem do informacji niejawnych. Warto mieć na uwadze, że nawet najdokładniejszy test, który ma wykrywać coś, czego w danej grupie badanych jest mało, przyniesie większą liczbę wyników fałszywie pozytywnych niż ten sam test (a nawet test mniej dokładny) wykonany w analogicznej grupie osób, pośród których wielu ma poszukiwaną cechę.

Nie jest możliwe, aby perfekcyjnie przełożyć na liczby wszystkie okoliczności, które przemawiają za daną hipotezą. Dlatego nawet jeśli zostanie podjęta taka próba, będzie to jedynie szacunek przybliżony. Dla przykładu wyobraźmy sobie, że na bezludną wyspę przypląwa na tratwie trzech rozbitków, z których jeden zostaje zamordowany wskutek zadania ciosu nożem. Żaden z pozostałych przybyszów nie przyznaje się do winy. Przy założeniu, że nie ujawniono okoliczności, które przemawiałyby za określoną wersją osobową, to przed badaniem dla obu badanych istnieją równe szanse winy (1:1) – wskaźnik bazowy wynosi więc 0,5. Generalnie w badaniach poligraficznych przyjmuje się taki właśnie wskaźnik na starcie diagnozowania wiarygodności. Jest to jedyne założenie, które można sprawiedliwie

Credibility Assessment: A Journal of Science and Field Practice” 2017, t. 46, nr 2, s. 115; tenże, *Multinomial Cutscores for Bayesian Analysis with ESS and Three-Position Scores of Comparison Question Polygraph Tests*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2020, t. 49, nr 1, s. 61–72.

zastosować w codziennej praktyce, ale też czysto teoretyczne, wręcz utopijne. W realnych warunkach występuje wiele różnych okoliczności, które będą wpływały na to, jak bardzo dana hipoteza jest prawdopodobna. W omawianym przykładzie byłyby to na przykład ewentualne ślady zbrodni (inne niż ślady pamięciowe).

Jeżeli osoba poddana testom z wykorzystaniem poligrafu reaguje na pytania krytyczne w sposób znaczący, tak jak zwykle robią to osoby, które kłamią lub zatajają, a inne przyczyny są mało prawdopodobne – to w drodze wniosku redukcyjnego uznaje się, że rzeczywiście mówi nieprawdę. Gdy przyjmie się, że jeden z rozbitków kłamie, a inne dowody nie dostarczają istotnych argumentów przemawiających za jego niewinnością i nie obciążają drugiego, to ten pierwszy najpewniej zostanie uznany za winnego. Może to być podejście zawodne – prawdziwe następstwo prowadzi niekiedy do fałszywej racji (odwrotnie niż w dedukcji, gdzie wcześniejszy zbiór prawdziwych przesłanek zawsze prowadzi do niekwestionowanych wniosków). Jednak w życiu codziennym znaczna część naszego rozumowania wygląda podobnie, subiektywnie, chociaż może dosłownie nie przeliczamy za każdym razem siły argumentów. Potrzebujemy rozstrzygnięć, decyzji, nawet gdy opieramy się na informacjach w jakiejś mierze niepewnych. Sąd, który ustala określony stan faktyczny, daje wyraz swojemu dostatecznemu przekonaniu, że inne wersje są nieprawdopodobne względem tej ostatecznie przyjętej. Gdyby była wymagana pewność absolutna, to co prawda nigdy nie skrzywdzono by niewinnego (faktyczna liczba pomyłek sądowych i tak pozostaje niewiadomą i nie zostanie poznana), ale też żaden przestępca nie zostałby ukarany.

Ograniczenia wynikające z wielokrotnego testowania

W przepisach dotyczących poszczególnych instytucji w różny sposób podchodzi się do zagadnienia powtarzania czynności w ramach postępowania kwalifikacyjnego. Dla części podmiotów nie ma konkretnych zapisów, które odnosiłyby się w tym aspekcie do badań poligraficznych. Z kolei np. w § 14 *Rozporządzenia Ministra Finansów z dnia 28 marca 2018 r. w sprawie przeprowadzania badania psychofizjologicznego, testu sprawności fizycznej oraz badania psychologicznego funkcjonariuszy Służby Celno-Skarbowej* można przeczytać, że (...) jeżeli wynik przeprowadzonego badania psychofizjologicznego jest niejednoznaczny i nie ma możliwości sporządzenia na jego podstawie sprawozdania, (...) badanie można powtórzyć jeden raz, na podstawie tego samego wniosku, w terminie nieprzekraczającym 30 dni od dnia uzyskania tego wyniku. Wydaje się również, że nie ma formalnych przeszkód dla zlecenia kolejnego badania przy sporządzeniu odrębnego wniosku.

Na podstawie wymiany doświadczeń między poligraferami na seminariach krajowych i międzynarodowych można wnioskować, że w części instytucji powtarza się wyłącznie testy zakończone nierozstrzygniętym rezultatem oraz w razie pojawienia się nowych okoliczności na temat kandydata, które wymagają weryfikacji za pomocą badania poligraficznego. W innych praktykuje się ponowne badanie w ramach „drugiej szansy”, czyli wykonuje się je również wtedy, gdy pierwsze przyniosło niekorzystne wskazania w testach, a jednocześnie w jego toku nie uzyskano od osoby badanej oświadczeń, które jednostka prowadząca postępowanie kwalifikacyjne uznalaby po lekturze sprawozdania za dyskwalifikujące. Wydaje się, że taka powtórka nie wszędzie jest dopuszczalna. Na przykład w przepisie zawartym w § 5 ust. 5 rozporządzenia w sprawie postępowania kwalifikacyjnego do służby w ABW wprowadzonym w nowelizacji z 31 stycznia 2022 r. (*Rozporządzenie Prezesa Rady Ministrów z dnia 31 stycznia 2022 r. zmieniające rozporządzenie w sprawie wzoru kwestionariusza osobowego oraz szczegółowych zasad i trybu przeprowadzania postępowania kwalifikacyjnego wobec kandydatów do służby w Agencji Bezpieczeństwa Wewnętrznego*) wskazano, że powtórzenie jednego z etapów postępowania (w tym badania psychofizjologicznego) jest możliwe (...) w przypadku pojawienia się nowych okoliczności dotyczących kandydata, mogących wpłynąć na wynik (...), w celu wyjaśnienia zaistniałych wątpliwości. Bez wskazania nowych okoliczności (wśród których mogłyby być jakieś dodatkowe oświadczenia kandydata lub informacje powzięte na jego temat z innych źródeł) powtórne badanie należy zatem uznać za bezprzedmiotowe. Zignorowanie tego wymogu byłoby nie tylko sprzeczne z prawem, lecz także wątpliwe ze względów metodologicznych, o czym w dalszej części artykułu.

Bez wątpienia każdy przypadek wymaga indywidualnego podejścia, podlega też wielowymiarowej analizie na kilku szczeblach hierarchicznej struktury, na której szczycie znajduje się szef danego urzędu, instytucji. Nie da się do wszystkich przyłożyć jednej miary, ponieważ każdy człowiek z natury rzeczy jest inny. Zrozumiałe i usprawiedliwione, a nawet w pewnych sytuacjach warte wsparcia, jest niejednakowe podejście do kandydatów do służby oraz do funkcjonariuszy i pracowników już zatrudnionych (i często mających duży staż). Można to wyjaśnić na przykładzie. W obu przypadkach zostały wykonane przesiewowe badania poligraficzne (u kandydata było to typowe badanie wykonywane przed przyjęciem, a u funkcjonariusza – badanie kontrolne). Jedno i drugie dostarczyło wątpliwości związanych z jakimś zagadnieniem (nawet tożsamym, np. dotyczącym zażywania nielegalnych substancji psychoaktywnych). Decydenci rozważający dalsze kroki mają po jednej stronie hipotetycznego kandydata, po drugiej – funkcjonariusza. W przypadku tego pierwszego ewentualne odrzucenie aplikacji oznacza niepodjęcie ryzyka, które nie musi się zmaterializować, ale też konsekwencje, straty są niewielkie. Jedyne zainwestowane nakłady pracy i środki finansowe wiążą się z przeprowadzeniem

wcześniejszych etapów postępowania kwalifikacyjnego. Sam kandydat, poza własnym czasem przeznaczonym na odbycie kilku spotkań i dojazdu, również nie zdążył jeszcze wiele zainwestować. Odmiennie przedstawia się sytuacja funkcjonariusza, na którego przeszkolenie przeznaczono określone nakłady. Jest to człowiek, który ma już konkretną wiedzę, zna tajemnice, wniósł też zapewne mniejszy lub większy wkład w rozwój instytucji. U kandydata należy ocenić, czy dana okoliczność zasygnalizowana po badaniu poligraficznym może przełożyć się negatywnie na jego funkcjonowanie w miejscu pracy albo służby. U osoby zatrudnionej – czy może ona powodować poważne problemy, ale przede wszystkim – czy dotąd je powodowała. Jeśli nie, to czy osoba zatrudniona rokuje poprawę i jest możliwe zastosowanie środków wspierającego bądź sankcyjnego oddziaływania. Dla lepszego zobrazowania tych różnic można odwołać się do relacji partnerskich. Gdy na pierwszym spotkaniu jednej z osób nie spodoba się zachowanie drugiej strony, jest bardzo prawdopodobne, że będzie to zarazem spotkanie ostatnie. Natomiast w dłużej trwającym związku, zacieśnionym więzami formalnymi i wspólnymi doświadczeniami, zanim nastąpi pożegnanie, zwykle jest podejmowana próba zwrócenia uwagi na problem lub wymuszenia jakiejś zmiany. Jeśli się to udaje, następuje wzmocnienie relacji. Powtarzanie się zachowań niepożądanych powoduje jednak, że motywacja do poszukiwania porozumienia spada i pojawia się przestrzeń dla bardziej radykalnych rozwiązań.

To zróżnicowane podejście widać na przykładzie kryteriów orzekania wyznaczonych komisjom lekarskim. Zapewne wzięto w nich pod uwagę to, że choroby mogą zostać nabyte w trakcie służby, w tym również w związku ze służbą. W *Rozporządzeniu Prezesa Rady Ministrów z dnia 15 kwietnia 2003 r. w sprawie oceny zdolności fizycznej i psychicznej do służby w Agencji Bezpieczeństwa Wewnętrznego* znajduje się zapis, że „jednostronny niedosłuch w paśmie częstotliwości niskich i średnich” dyskwalifikuje kandydata, ale nie dyskwalifikuje funkcjonariusza przewidzianego do dalszej służby. Podobnie jest między innymi z „zaburzeniami nerwowymi w trakcie leczenia, rokującymi poprawę” czy z „używaniem szkodliwym alkoholu i środków odurzających”.

Różne instytucje, tak w Polsce, jak na świecie, w odmienny sposób podchodzą do ponownego testowania z wykorzystaniem poligrafu przy niezmiennym zakresie merytorycznym badania. Decyzje z tym związane mają charakter rutynowy, proceduralny (bez względu na wynik pierwszego badania) albo doraźny (w zależności od wyniku – wskazującego na szczerść lub nieszczerść kandydata, ewentualnie brak jednoznacznego rozstrzygnięcia). W pierwszym przypadku badanie poligraficzne inicjuje i wieńczy proces rekrutacji. W drugim z reguły ponownie wykonuje się badania nierozstrzygnięte – bez większych kontrowersji. W części podmiotów powtarza się – niejako dla pewności – testy zakończone wynikiem NDI (ang. *no deception*

indicated – brak wskazania na wprowadzanie w błąd), a przeciwnie rezultaty kończą procedurę. W innych odwrotnie – powtórne czynności zleca się w razie wystąpienia wyniku DI (ang. *deception indicated* – wskazanie na wprowadzanie w błąd), a wynik korzystny dla badanego pozwala na kontynuowanie postępowania. Jakie są implikacje powyższych rozwiązań? Każde z nich ma swoje zalety i wady. W zależności od specyfiki danej instytucji jedno mogą być bardziej optymalne, inne mniej.

Przeciętne przesiewowe badanie poligraficzne przeprowadzane w ramach postępowań rekrutacyjnych polega (pomijając inne etapy badania, jak wywiad przedtestowy) na wykonaniu testu wieloprogowego (zawierającego pytania związane z wykazem kluczowych zagadnień przy ocenie predyspozycji kandydata do pracy lub służby). Po stwierdzeniu znaczących zmian reakcji fizjologicznych poligrafer zwykle decyduje się na drugi test (wówczas już podczas pierwszego badania *de facto* jest przeprowadzane powtórne testowanie). Ten drugi test wybiera się najczęściej z grupy testów diagnostycznych (jednowątkowych), o wyższej dokładności, swoistości (nadal jednak – ze względu na szeroki zakres krytycznego zagadnienia, np. popełnienie przestępstwa, bez wskazywania kategorii, tym bardziej bez odniesienia się do czynu w konkretnym czasie i miejscu – badanie utrzymuje swój przesiewowy charakter).

Zaleca się, aby test powtórny był odmienny od pierwszego (w świecie idealnym inny powinien być nie tylko test, lecz także urządzenie, metoda analityczna oraz sam badający). Ma to zapobiec sytuacji, w której ta sama nielosowa przyczyna mogąca odpowiadać poprzednio za błąd zaistnieje ponownie w niezmiennych warunkach⁶. Dążenie do zapewnienia obiektywizmu i odporności na wpływ wcześniejszych informacji powinno też skłaniać do tego, aby kolejnego poligrafera nie zaznajamiać z wynikami wcześniejszych testów. Pozostaje praktyczny problem odpowiedniego przygotowania się do takiego badania. Czasami nie ma możliwości zastąpienia osoby, która przeprowadziła poprzednie badanie. Wówczas najlepiej jest, aby badający ściśle trzymał się ustrukturyzowanej procedury prowadzenia wywiadu, stosował takie techniki, które mniej zależą od osobistych właściwości poligrafera, np. pytania kontrolne typu DLC (ang. *directed lie comparisons* – pytania porównawcze z poleconym kłamstwem), wspomagał się komputerową, algorytmiczną analizą zarejestrowanych zapisów, poddał się kontroli jakości z uwzględnieniem ślepej interpretacji wykresów.

Po uzyskaniu wyniku testu wskazującego na nieszczerłość niekiedy przeprowadza się jeszcze testy pomocnicze, naprowadzające na bardziej szczegółowe aspekty

⁶ Por. R. Nelson, F. Turner, *Bayesian Probabilities of Deception and Truth-telling for Single and Repeated Polygraph Examinations*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2017, t. 46, nr 1, s. 62.

ujawnionego problemu. Trzeba zaznaczyć, że w wielu instytucjach, zwłaszcza amerykańskich, badanie ogranicza się wyłącznie do jednego testu przesiewowego, a drugi wykonuje się wyjątkowo. Może się tak zdarzyć, jeżeli w toku tego badania kandydat złoży nowe oświadczenie uzasadniające weryfikację uzupełniającą (na zasadzie – czy to rzeczywiście wszystko, co należało dodać, czy też problem pozostaje i tkwi w jeszcze innym wspomnieniu lub wspomnieniach, których badany nie przekazał).

Podczas badania przesiewowego nie jest wiadome, ile wynosi wstępne prawdopodobieństwo występowania danej cechy w badanej populacji. Z tego względu w badaniach poligraficznych przyjmuje się standardowo szanse 1:1 (jest tak samo prawdopodobne, że badany będzie szczerzy i będzie nieszczerzy). Nie wiadomo z góry (przy robieniu szacunków nie jesteśmy jednak zupełnie bezsilni⁷), jaka część populacji biorącej udział w rekrutacji ukrywa np. fakt zażywania narkotyków albo kontakty z grupami podwyższonego ryzyka. Intuicyjnie można jednak założyć, że tych pierwszych będzie więcej niż drugich. W konsekwencji – jeżeli uzyska się pozytywne (czyli te złe) wskazania testów, to prawdopodobieństwo, że wskazanie jest prawdziwe, będzie wyższe w odniesieniu do narkotyków niż np. szpiegostwa. Można to pokazać na liczbach. Prawdopodobieństwo występowania u badanego cechy zasygnalizowanej w teście (np. ukrywania czegoś w kwestii narkotyków) można obliczyć następująco (czyniąc hipotetyczne założenie, że osobiste doświadczenia z narkotykami miało 20% populacji):

$$\begin{aligned} & \text{Prawdopodobieństwo a posteriori po teście negatywnym („NSR”/„NDI”):} \\ & = \frac{\text{prawdopodobieństwo wstępne szczerości x wskaźnik swoistości testu}}{\text{(prawdopodobieństwo wstępne x swoistość testu)} + [(1 - \text{wstępne prawdopodobieństwo}) \times \text{wskaźnik wyników fałszywych negatywnych}]} \end{aligned}$$

Jeśli test przesiewowy został przeprowadzony w standaryzowanym formacie DLST (ang. Directed Lie Screening Test – Test przesiewowy z ukierunkowanym kłamstwem) i do analizy danych wykorzystano system ESS (ang. Empirical Scoring System – Empiryczny system oceniania), to wiadomo, że jego czułość wynosi 0,809, a wskaźnik wyników fałszywych pozytywnych (FP) – 0,146⁸. Analiza zapisów na poligramach kończy się opinią: SR (ang. *significant responses* – znaczące reakcje) w konsekwencji sposobu reagowania badanego na pytanie o ukrywanie

⁷ Przybliżone liczby można niekiedy zaczerpnąć z różnych źródeł informacji – danych historycznych, badań naukowych, ankiet, statystyk policyjnych, sądowych i innych. Dla porównania – ciekawą metodą jest badanie stężenia wirusów w ściekach jako próba monitorowania rzeczywistej sytuacji epidemiologicznej.

⁸ American Polygraph Association, *Meta-Analytic Survey...*, s. 244.

informacji związanych z narkotykami. Gdy pod powyższy wzór zostaną podstawione dostępne liczby, to uzyskuje się:

$$\text{prawdopodobieństwo a posteriori} = \frac{0,20 \times 0,809}{(0,20 \times 0,809) + [(1 - 0,20) \times 0,146]} = 0,58$$

W sytuacji opisanej w przykładzie po jednym teście przesiewowym oszacowano, że prawdopodobieństwo ukrywania przez hipotetycznego badanego jakiegoś problemu z narkotykami wynosi 58%. Czy to jest mało? Pozornie. Chociaż jeśli dotąd przy wyciąganiu wniosków wzięto po uwagę tylko średnią dokładność testu DLST wynoszącą 85,8%, to różnica jest widoczna.

Z jeszcze mniejszą pewnością będziemy mieli do czynienia w sytuacji, gdy „alarm” z testu przesiewowego zaistnieje przy incydencie, którego realne występowanie w populacji osób badanych jest jeszcze rzadsze niż w wypadku narkotyków wykorzystanych w przykładzie. Niech to będą kontakty z obcymi służbami specjalnymi. Można subiektywnie przyjąć, że w takie uwikłał się 1% populacji kandydatów i został wykonany ten sam test (DLST), zakończony pozytywnym wskazaniem. Po podstawieniu danych do wzoru można stwierdzić, że po przeprowadzonym teście prawdopodobieństwo, że badany jest szpiegiem, wynosi zaledwie 5,3%. Poprzedzenie na tym etapie spowodowałoby, że większość takich wskazań byłaby realnie fałszywa. Dlatego między innymi w okresie pandemii COVID-19 raczej mieli ci, którzy uważali, że powszechne testowanie całej populacji nie ma sensu, ponieważ wiele osób trafiłoby na izolację bez powodu, zwłaszcza w okresie, kiedy zakładany wskaźnik zachorowań wśród ogółu ludności był już relatywnie niewielki. Testowanie wszystkich byłoby racjonalnym działaniem najwyżej podczas szczytu pandemii. Właściwym podejściem było kierowanie ludzi na dodatkowe testy diagnostyczne, genetyczne typu PCR, po pozytywnym wskazaniu szybkiego testu przesiewowego, antygenowego. Podobny tok postępowania ma zastosowanie w badaniach poligraficznych.

Wskazania testów przesiewowych słusznie traktuje się jako powód do dalszego testowania. Jest zasadne przeprowadzenie po takim pierwszym teście kolejnego, najlepiej dokładniejszego, z wachlarza testów diagnostycznych. Na przykładzie opisanego w artykule hipotetycznego badania można zaobserwować, jak zmieni się sytuacja w kwestii narkotyków po wykonaniu kolejnego testu. Przed badaniem oszacowano, że prawdopodobieństwo kontaktu badanego z narkotykami wynosi 20%. Po teście przesiewowym to prawdopodobieństwo wzrosło do 58%. Zostaje przeprowadzony dodatkowy test – diagnostyczny, np. w standaryzowanym formacie You-Phase (Backster Bi-Zone), o średniej dokładności 90,4%. Do analizy

danych testowych ponownie zostanie wykorzystany system ESS. Test charakteryzuje się wówczas czułością wynoszącą 0,845 i wskaźnikiem wyników fałszywych pozytywnych na poziomie 0,138⁹. Analiza kończy się rozstrzygnięciem DI. Należy podstawić dane do wzoru na prawdopodobieństwo a posteriori (prawdopodobieństwo a priori po pierwszym teście, a przed drugim wynosi 0,58).

$$\text{prawdopodobieństwo po teście nr 2} = \frac{0,58 \times 0,845}{(0,58 \times 0,845) + [(1 - 0,58) \times 0,138]} = 0,89$$

Na tym etapie stopień pewności, że problem naświetlony po pierwszym teście przesiewowym rzeczywiście występuje, jest już znacznie większy. Po teście nr 2 prawdopodobieństwo wzrosło do 89%.

Statystyczne właściwości danego testu (m.in. czułość, swoistość, poziom błędu decyzyjnego) nie zawsze bezpośrednio przekładają się na wnioski dotyczące rzeczywistej siły relacji między dwiema zmiennymi w populacji statystycznej. Jak zauważają Raymond Nelson i Finley Turner, teoremat Bayesa pomaga skuteczniej i trafniej opisywać relacje z realnego świata, zważywszy na wstępne założenia, jakie zostały poczynione względem jakiegoś problemu badawczego¹⁰.

Dla porządku należy przedstawić jeszcze wariant testu przesiewowego, w którym nie zarejestrowano znaczących zmian reakcji fizjologicznych, np. z pytaniami o przestępstwa i istnienie materiałów nacisku na badanego. Przyjmijmy, że w całym procesie badawczym zostanie wykorzystany wyłącznie jeden test w standaryzowanym formacie USAF MGQT i analiza danych zostanie przeprowadzona zgodnie z regułami „systemu federalnego”, w skali 7-pozycyjnej. W tych warunkach swoistość testu wynosi 0,538. Z kolei odsetek wyników fałszywych negatywnych (FN) wynosi 0,079¹¹. Uzyskany rezultat to NSR (ang. *no significant responses* – brak znaczących reakcji). Nie są znane dane, które pozwoliłyby adekwatnie dostosować prawdopodobieństwo wstępne (a priori), więc zostanie przyjęte standardowe (0,5). Należy zmodyfikować odpowiednio wzór:

⁹ Tamże, s. 240.

¹⁰ Zob. R. Nelson, F. Turner, *Bayesian Probabilities of Deception...*, s. 53–80.

¹¹ American Polygraph Association, *Meta-Analytic Survey...*, s. 244.

$$\begin{aligned} & \text{Prawdopodobieństwo a posteriori po teście negatywnym („NSR”/„NDI”):} \\ = & \frac{\text{prawdopodobieństwo wstępne szczerości} \times \text{wskaźnik swoistości testu}}{(\text{prawdopodobieństwo wstępne} \times \text{swoistość testu})} \\ & + [(1 - \text{wstępne prawdopodobieństwo}) \times \text{wskaźnik wyników fałszywych negatywnych}] \end{aligned}$$

Następnie podstawić dane:

$$\text{prawdopodobieństwo a posteriori} = \frac{0,5 \times 0,538}{(0,5 \times 0,538) + [(1 - 0,5) \times 0,079]} = 0,87$$

Po tym teście realne prawdopodobieństwo, że badany nie kłamał, gdy twierdził, że nie dopuścił się przestępstw i nie ukrywa okoliczności stanowiących potencjalny materiał nacisku, wynosi 87%. Przed testem przyjęto, że hipoteza o prawdziwości jest prawdopodobna na 50%, a po jego wykonaniu to prawdopodobieństwo zwiększyło się o 37 punktów procentowych.

Zachęcające wyniki dwuetapowego modelu testowania opisanego wyżej w ramach hipotetycznego badania z ujawnieniem problemu związanego z narkotykami mogą skłonić do zastanowienia się, czy wykonywanie większej liczby testów na ten sam temat nie byłoby słusznym posunięciem, skoro prawdopodobieństwo trafnej diagnozy tak bardzo wzrosło po drugim teście. Otóż sprawa nie jest taka prosta, korzyści okazują się tylko pozorne, a ryzyko nieprawidłowych rozstrzygnięć niebezpiecznie rośnie. Ponadto dotąd były rozważane jedynie wskazania, które są ze sobą spójne (w kierunku wskazującym na nieszczerłość). Tymczasem może zdarzyć się sytuacja, gdy wyniki dwóch kolejnych testów (nieważne, czy wykonanych w ramach jednego badania czy podczas dwóch odrębnych spotkań) będą rozbieżne (zwykle chodzi o rozbieżność DI/NDI). Jeśli osoba badana nie przekaze między tymi testami istotnych oświadczeń uzupełniających (np. przyznając się do jakiegoś incydentu), nie będzie to przypadek częsty. Niemniej jednak jest to z różnych przyczyn możliwe (te przyczyny nie zostaną w tym miejscu szerzej omówione, gdyż jest to temat na odrębne opracowanie; na potrzeby niniejszych analiz zostały wzięte pod uwagę wyłącznie właściwości samych testów).

Kiedy badanie poligraficzne jest ostatnim lub jednym z ostatnich etapów postępowania kwalifikacyjnego, kandydat ma za sobą zwykle rozmowę wstępną, badanie psychologiczne i inne testy – np. wiedzy, sprawnościowe, być może uzyskał już opinię właściwej komisji lekarskiej, podlegał sprawdzeniom (w tym kartotecznym czy w ramach wywiadu środowiskowego). Komórki HR (zwłaszcza w obliczu zwiększonych potrzeb kadrowych) będą więc niechętnie poprzestaniu na jednym

niekorzystnym wyniku badania poligraficznego, który poniekąd niweczyłby dotychczasowy wysiłek. Zwiększa to motywację do tego, aby przeprowadzić badanie „drugiej szansy”. Bywa, że w takiej sytuacji kandydat uzupełni informacje na swój temat o takie, których nie przekazał podczas pierwszego badania, i to umożliwi pomyślne zakończenie procesu rekrutacji. Zdarza się także, że nowe oświadczenia się nie pojawiają i kolejny test kończy się określonym wynikiem – częściej spójnym z wynikiem testu podstawowego, ale niekiedy przeciwnym. Który wynik jest wówczas bardziej wiarygodny? Może w sytuacji zaistnienia rozbieżności za konieczne należałoby uznać trzecie badanie? Wszystkie opisane scenariusze budzą kontrowersje – włącznie z pierwszym (zakładającym uzupełniające informacje), który teoretycznie jest najbardziej klarowny. U kandydata, który skorygował swoje stanowisko (czyli przyznał się w końcu do czegoś, co ukrywał na początku), może bowiem pojawić się przeświadczenie, że warto za pierwszym razem zaryzykować kłamstwo, przemilczeć coś, ponieważ instytucja to toleruje i będzie możliwość ewentualnej poprawki, a przy łucie szczęścia – uda się za pierwszym razem. Z kolei pozostałe scenariusze rodzą określone konsekwencje związane ze specyfiką testów przesiewowych i nieuchronnością pojawienia się jakiejś liczby wyników fałszywych – zarówno pozytywnych, jak i negatywnych.

Z badań przeprowadzonych przez Wiliama Yankeego i Douglasa Grimsleya wynika, że z jednej strony co prawda nie stwierdzono istotnych statystycznie różnic między dokładnością testu wykonanego podczas pierwszego badania i powtórnego testu w tej samej sprawie (co dobrze świadczy o rzetelności testu), lecz z drugiej strony w przypadku badanych występujących w roli wprowadzających w błąd zaobserwowano wzrost liczby wyników fałszywych negatywnych, a ogólna dokładność (po wyłączeniu rezultatów nierozstrzygniętych) spadła z 87% do 76% (być może jest to konsekwencja zjawiska habituacji, czyli stopniowego przyzwyczajania się do wielokrotnie powtarzanych bodźców testowych, być może zastosowania jakichś psychologicznych mechanizmów obronnych – nie ma w tej kwestii pewnej odpowiedzi). Sytuacja praktycznie w ogóle nie zmieniła się natomiast w grupie osób szczerych (odpowiednio: 100% i 96% poprawnych wskazań)¹². Na próżno szukać badań empirycznych, które pokazywałyby odwrotną tendencję. Daje to więc do myślenia na temat tego, który wynik uznać za bardziej wiarygodny w sytuacji rozbieżności i przekształcenia z DI w NDI (pomijając przy tym, czy między pierwszym a drugim badaniem nastąpiła zmiana kontekstu badawczego z powodu złożenia przez kandydata nowych oświadczeń).

¹² W. Yankee, D. Grimsley, *Test and Retest Accuracy of Psychophysiological Detection of Deception Test*, „Polygraph” 2000, t. 29, nr 4, s. 289–298.

Zobaczmy, jak statystycznie może wyglądać przebieg procedury rekrutacyjnej do jakiejś instytucji na przykład dla ogółu kandydatów w skali roku. Hipotetyczny scenariusz przewiduje, że w postępowaniu kwalifikacyjnym bierze udział 750 kandydatów. Według subiektywnych, apriorycznych kalkulacji rekruterów 20% tej populacji ma cechy niepożądane. Nie licząc testu demonstracyjnego, od którego zaczyna się zasadniczy etap każdego badania poligraficznego, wszystkim wykonano test przesiewowy we wspomnianym już formacie DLST i z wykorzystaniem systemu analizy ESS, z następującymi parametrami: czułość – 0,809; FP – 0,146; swoistość – 0,751; FN – 0,112. Jakich liczb należy się spodziewać? W kontekście powtórnego testowania w polu uwagi są głównie wyniki nierozstrzygnięte (ang. *inconclusives* – INC) i wskazujące na nieszczerość (ściślej SR – znaczące reakcje). Po uwzględnieniu wymienionych parametrów otrzyma się następujące działania:

- czułość testu (0,809) x prawdopodobieństwo wstępne nie szczerości (0,2) x liczba badanych osób (750) = 121 wyników prawdziwie pozytywnych;
- swoistość testu (0,751) x prawdopodobieństwo wstępne szczerości (0,8) x liczba osób (750) = 451 wyników prawdziwie negatywnych;
- FN (0,112) x prawdopodobieństwo wstępne nie szczerości (0,2) x liczba badanych osób (750) = 17 wyników fałszywie negatywnych;
- FP (0,146) x prawdopodobieństwo wstępne szczerości (0,8) x liczba badanych osób (750) = 88 wyników fałszywie pozytywnych;
- pozostałe testy kwalifikuje się do zbioru nierozstrzygniętych (750 – 677 = 73).

Gdyby postępowanie kwalifikacyjne zostało zakończone na tym etapie, to pracodawca podziękowałby 88 (11,7%) kandydatom, którym mógłby dać szansę. Jednocześnie przyjąłby 17 (2,3%) takich, których wolałby nie zatrudniać. Z powodu niejednoznacznego wyniku na kolejny test musiałyby zostać zaproszone 73 osoby (9,7%). Pozostałe decyzje kadrowe – dotyczące 572 z 662 osób (86,4%) – byłyby trafne.

W przesiewowych badaniach poligraficznych stosuje się zazwyczaj podejście „sukcesywnego pokonywania przeszkód” (ang. *successive hurdles*). Negatywne wyniki testów (czyli te, które nie wskazują żadnego problemu – NSR, NDI) zamykają ten etap i stanowią dla kandydata przepustkę do kolejnego. Pozytywne wyniki wymagają natomiast przeprowadzenia dodatkowych testów lub innych czynności weryfikujących, zanim zostaną uznane za podstawę jakiegoś działania czy decyzji. Powtarza się też, co oczywiste, badania nierozstrzygnięte.

Testy (zarówno pierwsze, jak i kolejne) mogą przynosić wyniki trafne bądź nietrafne, spójne ze sobą i rozbieżne. Jeśli przyjmie się w przybliżeniu, że średnia dokładność testu przesiewowego wynosi 85%, to uzyskuje się następujące warianty wyników:

- dwa niezależne testy z prawidłowymi i zgodnymi wynikami: $0,85 \times 0,85 = 0,72$;
- dwa niezależne testy z nietrafnymi i spójnymi wynikami: $0,15 \times 0,15 = 0,02$;
- suma spodziewanej zgodności dwóch niezależnych testów wyniesie: $0,72 + 0,02 = 0,74$ i odwrotnie – w 26% przypadków trzeba się liczyć z potencjalną rozbieżnością rezultatów;
- łączna dokładność dwóch niezależnych, spójnych ze sobą testów: $0,72 / 0,74 = 0,97$.

Rozszerzmy jeszcze opisane warianty o kasus polegający na wykonaniu podczas pierwszego badania dwóch testów (przesiewowego i diagnostycznego, dokładnych na odpowiednio: 85% i 90%)¹³, a następnie wykonaniu oczekiwanego przez zleceniodawcę powtórnego badania, w ramach którego inny poligrafer przeprowadzi trzeci test (diagnostyczny, o dokładności 90%):

- trzy testy z prawidłowymi i zgodnymi wynikami: $0,85 \times 0,90 \times 0,90 = 0,69$;
- trzy testy z nietrafnymi i spójnymi wynikami: $0,15 \times 0,10 \times 0,10 = 0,0015$;
- suma spodziewanej zgodności trzech testów wyniesie: $0,69 + 0,0015 = 0,6915$; z kolei w ok. 31% przypadków mogą wystąpić rozbieżne rezultaty. Jak widać, z każdym kolejnym testem prawdopodobieństwo wystąpienia odmiennego wyniku od dotychczasowych rośnie, a w konsekwencji wzrasta również ryzyko błędu, jeśli za podstawę rozstrzygnięcia zostałby wzięty tylko ostatni test, przy którym nastąpiła zmiana;
- łączna dokładność trzech spójnych testów wyniesie: $0,69 / 0,6915 = 0,998$.

Wróćmy do opisanego przykładowego postępowania kwalifikacyjnego, w którym wzięło udział 750 kandydatów. Po pierwszym teście 209 osób uzyskało wyniki pozytywne i skierowano je na powtórny test. Wskaźnik PPV (ang. *positive predictive value* – wartość predykcyjna dodatnia), czyli ile wyników spośród pozytywnych jest prawdziwych, wynosi teraz 0,727. To oznacza, że założone przez rekruterów wstępne prawdopodobieństwo wprowadzania w błąd wzrosło po pierwszym teście w grupie z wynikami pozytywnymi z 20% do 72,7%. Wskaźnik NPV (ang. *negative predictive value* – wartość predykcyjna ujemna), określający odsetek prawdziwych wyników negatywnych, oszacowano natomiast na 0,273. Po zastosowaniu testu You-Phase / ESS otrzymuje się więc:

¹³ Należy zastrzec, że testy wykonywane przez tego samego badającego w ramach jednej czynności – nawet jeśli zostanie zmieniony rodzaj testu – nie są od siebie w pełni niezależne. Część zmiennych, niepozostających bez wpływu na wyniki, współwystępuje. Zaprezentowany przykład stanowi pewne uproszczenie – tak jakbyśmy mieli do czynienia za każdym razem z niezależnymi od siebie testami. Szerzej na ten temat: R. Nelson, J. Kircher, M. Handler, *How to Calculate the Expected Agreement and the Combined Accuracy of Two Test Results*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2018, t. 47, nr 1, s. 18–25.

- 128 wyników prawdziwie pozytywnych = czułość testu (0,845) x PPV (0,727) x liczba badanych osób (209);
- 43 wyniki prawdziwie negatywne = swoistość testu (0,757) x NPV (0,273) x liczba badanych osób (209).

Nie uniknie się następujących wyników fałszywych:

- 20 fałszywie negatywnych = FN (0,134) x PPV (0,727) x liczba badanych osób (209);
- 8 fałszywie pozytywnych = FP (0,138) x NPV (0,273) x liczba badanych osób (209);
- pozostałe wyniki zakwalifikują się do zbioru nierozstrzygniętych (209 – 199 = 10).

Liczby pokazują, że druga tura testów umożliwiłaby skorygowanie fałszywie obciążających wyników u 43 osób (21% powtórnie badanych, zatem jest to niezły wynik), ale efektem ubocznym byłoby 20 kolejnych wyników fałszywych negatywnych. To niestety oznacza, że w przedstawionych warunkach niemal co trzecie badanie z wynikiem negatywnym jest fałszywe (32%), a w liczbach bezwzględnych do instytucji zostanie przez to wpuszczonych nawet więcej osób niepożądanych niż podczas pierwszych testów, ale wtedy wzięło w nich udział ponad 3,5 razy więcej kandydatów. W takiej sytuacji pozostaje mieć nadzieję, że inne „bezpieczniki” nie zawiodą. Jeśli jednak badanie poligraficzne jest umiejscowione niemal na końcu postępowania, to może ich być już niewiele.

Wnioski i rekomendacje

Powyższe rozważania mają skłaniać nie tyle do liczbowego szacowania wszystkiego, ile do prawidłowej interpretacji znaczenia wyników różnych testów (w tym poligraficznych), przypisywania im odpowiedniej wagi, dostosowania do tego procesów decyzyjnych i uświadomienia potencjalnych konsekwencji różnych działań, np. bilansu zysków i strat związanych z powtórным testowaniem.

Dokładność testowania rośnie tylko wtedy, gdy kolejne próby dają spójne rezultaty. Powtórne badania osób, dla których testy przebiegły niepomyślnie, pozwalają skorygować w dół wyłącznie liczbę wyników fałszywych pozytywnych, ale – ku rozczarowaniu wielu albo świadomej ignorancji – generalnie nie zwiększają dokładności testów przy wynikach odmiennych od tych uzyskanych za pierwszym razem. Podobne zjawisko występowałoby, gdyby powtarzano badania tylko tych kandydatów, którym udało się bez zastrzeżeń przejść próbę nr 1. Można byłoby wyeliminować z rekrutacji część niewiarygodnych osób, które przeszły przez pierwsze sito, ale całościowo też nie wyglądałoby to lepiej.

Nieuchronny wzrost liczby fałszywych wskazań negatywnych podczas potwórných badań stwarza palącą potrzebę, aby znaleźć rozwiązanie tego problemu. Na pewno sytuacja nie jest do zaakceptowania w instytucjach newralgicznych dla bezpieczeństwa kraju. Albo więc te sposoby się znajdują, albo należałoby całkowicie zrezygnować z „poprawkowych” badań i próbować weryfikować wątpliwości za pomocą innych narzędzi (o ile istnieją i są wystarczająco skuteczne). Jednym z rozwiązań jest dokładne wypytanie kandydata na temat krytycznych zagadnień, pomimo że udało mu się przejść test, ewentualnie – zdanie się na intuicję badającego w kwestii obrania kierunku dalszej rozmowy. Nelson i Turner konstatują, że takie podejście sprowadziłoby jednak testy poligraficzne do miana pseudonauki¹⁴. Ani naukowcy, ani praktycy nie mają zatem doskonałych recept.

Można odwołać się po raz kolejny do doświadczeń z pandemii COVID-19 i przypomnieć okresy, w których wyjście z izolacji musiało zostać poprzedzone dwoma kolejnymi negatywnymi wynikami testów. Schemat postępowania był taki: pozytywny wynik testu nr 1 (izolacja), negatywny wynik testu nr 2 (nadzieja na koniec izolacji), negatywny wynik testu nr 3 (status ozdrowieńca). Ta analogia jest bardzo adekwatna do badań poligraficznych, gdy zaczyna się od pozytywnego wyniku pierwszego testu i myśli się o kolejnym bądź kolejnych, które mogą dać wynik negatywny. Co więcej – dokładność testowania również wygląda podobnie, tak jak i podział na testy przesiewowe oraz diagnostyczne. Jeżeli zatem po wyniku SR/DI wystąpi NDI, należałoby rozważyć jeszcze jedno potwierdzenie, jeszcze jedno badanie.

Nowe oświadczenia kandydata w odniesieniu do krytycznego zagadnienia, które zagraża mu po pierwszym teście, stwarzają odmienne okoliczności badawcze. Mogą być realną przyczyną, wyjaśnieniem zarejestrowanych reakcji organizmu, ale nadal nie dają pewności, czy kandydat jeszcze czegoś nie ukrywa. Bez pojawienia się nowych okoliczności nie widać logicznych argumentów przemawiających za potwórnym badaniem.

Reasumując, dodatkowy test jest wysoce wskazany po pozytywnym (niekorzystnym) wyniku pierwszego testu przesiewowego. Jeśli jednak drugi test został wykonany już podczas pierwszego badania (a najczęściej tak bywa), to kolejne (oznaczające trzeci test i ewentualnie więcej) nie jest rekomendowane. Ryzyko wyników fałszywych negatywnych przeważa nad korzyściami płynącymi ze zmniejszenia wskazań fałszywych pozytywnych.

Skoro nie istnieje fenomen w postaci perfekcyjnego testu, to niemożliwe jest zapewnienie w równym stopniu ochrony interesów osoby aplikującej oraz instytucji

¹⁴ R. Nelson, F. Turner, *Bayesian Probabilities of Deception...*, s. 71.

przyjmującej (szczególnie w zakresie zapewnienia bezpieczeństwa wewnętrznego i ochrony informacji niejawnych). Dlatego trzeba się na coś zdecydować.

W różnych procedurach decyzyjnych priorytety kształtują się nieco inaczej. W procesie karnym obowiązuje zasada *in dubio pro reo* (art. 5 § 2 Kodeksu postępowania karnego). Ochronie informacji niejawnych przypisuje się w postępowaniach sprawdzających pierwszeństwo przed innymi prawnie chronionymi interesami (art. 24 ust. 4 *Ustawy z dnia 5 sierpnia 2010 r. o ochronie informacji niejawnych*).

Bezpieczeństwo wewnętrzne instytucji jest niewątpliwie nadrzędne względem interesu kandydata w postępowaniu kwalifikacyjnym do pracy lub służby. W innych postępowaniach kadrowych – wyjaśniających, dyscyplinarnych – priorytetem jest zapewnienie, by osoba niewinna nie poniosła odpowiedzialności, a w przypadku tych, którzy popełnili błąd, ważna jest ocena, czy rokują poprawę. Natomiast przy rekrutacji do pracy lub służby istotniejsze wydaje się, by niepożądana osoba (zwłaszcza pod względem aspektu jakiegoś relatywnie bardziej ryzykownego niż inne – np. kontrwywiadowczego) nie została przyjęta, nawet kosztem pewnego (niewielkiego) odsetka osób nietrafnie ocenionych, ale jeszcze bez poniesienia dużych nakładów i podjęcia wzajemnych zobowiązań na linii pracodawca–pracownik. Nierozważne „przepychanie” takich kandydatów przez zlecenie kolejnych badań aż do momentu osiągnięcia jednego korzystnego rezultatu testu (z coraz większym prawdopodobieństwem uzyskania wyniku fałszywego negatywnego) skutkowałoby wchodzeniem w struktury danej instytucji osób, które w skrajnych okolicznościach mogłyby stanowić zagrożenie bezpieczeństwa państwa.

Bibliografia

Addendum to the 2011 Meta-analytic Survey – the Utah Four-Question Test (“Raskin Technique”)/ ESS, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2020, t. 49, nr 2, s. 73–81.

American Polygraph Association, *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, „Polygraph” 2011, t. 40, nr 4.

Eggleston R., *Sixth Wilfred Fullagar Memorial Lecture: Beyond reasonable doubt*, „Monash University Law Review” 1977, t. 4, nr 1, s. 1–2.

Handler M. i in., *Integration of Pre-Employment Polygraph Screening into the Police Selection Process*, „Journal of Police and Criminal Psychology” 2009, t. 24, nr 2, s. 69–86.

Nelson R., *Multinomial Cutscores for Bayesian Analysis with ESS and Three-Position Scores of Comparison Question Polygraph Tests*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2020, t. 49, nr 1, s. 61–72.

Nelson R., *Multinomial Reference Distributions for the Empirical Scoring System*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2017, t. 46, nr 2, s. 81–115.

Nelson R. i in., *Using the Empirical Scoring System*, „Polygraph” 2011, t. 40, nr 2, s. 67–78.

Nelson R., Kircher J., Handler M., *How to Calculate the Expected Agreement and the Combined Accuracy of Two Test Results*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2018, t. 47, nr 1, s. 18–25.

Nelson R., Turner F., *Bayesian Probabilities of Deception and Truth-telling for Single and Repeated Polygraph Examinations*, „Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice” 2017, t. 46, nr 1, s. 53–80.

Yankee W., Grimsley D., *Test and Retest Accuracy of Psychophysiological Detection of Deception Test*, „Polygraph” 2000, t. 29, nr 4, s. 289–298.

Akty prawne

Ustawa z dnia 8 grudnia 2017 r. o Służbie Ochrony Państwa (t.j. DzU z 2023 r. poz. 66, ze zm.).

Ustawa z dnia 16 listopada 2016 r. o Krajowej Administracji Skarbowej (t.j. DzU z 2022 r. poz. 813, ze zm.).

Ustawa z dnia 5 sierpnia 2010 r. o ochronie informacji niejawnych (t.j. DzU z 2019 r. poz. 742, ze zm.).

Ustawa z dnia 9 czerwca 2006 r. o Centralnym Biurze Antykorupcyjnym (t.j. DzU z 2022 r. poz. 1900, ze zm.).

Ustawa z dnia 24 sierpnia 2001 r. o Żandarmerii Wojskowej i wojskowych organach porządkowych (t.j. DzU z 2021 r. poz. 1124, ze zm.).

Ustawa z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego (t.j. DzU z 2022 r. poz. 1375, ze zm.).

Ustawa z dnia 12 października 1990 r. o Straży Granicznej (t.j. DzU z 2022 r. poz. 1061, ze zm.).

Rozporządzenie Prezesa Rady Ministrów z dnia 31 stycznia 2022 r. zmieniające rozporządzenie w sprawie wzoru kwestionariusza osobowego oraz szczegółowych zasad i trybu przeprowadzania postępowania kwalifikacyjnego wobec kandydatów do służby w Agencji Bezpieczeństwa Wewnętrznego (DzU z 2022 r. poz. 263).

Rozporządzenie Ministra Finansów z dnia 28 marca 2018 r. w sprawie przeprowadzania badania psychofizjologicznego, testu sprawności fizycznej oraz badania psychologicznego funkcjonariuszy Służby Celno-Skarbowej (t.j. DzU z 2022 r. poz. 379).

Rozporządzenie Prezesa Rady Ministrów z dnia 15 kwietnia 2003 r. w sprawie oceny zdolności fizycznej i psychicznej do służby w Agencji Bezpieczeństwa Wewnętrznego (t.j. DzU z 2014 r. poz. 242).

Rozporządzenie Prezesa Rady Ministrów z dnia 29 listopada 2002 r. w sprawie wzoru kwestionariusza osobowego oraz szczegółowych zasad i trybu przeprowadzania postępowania kwalifikacyjnego wobec kandydatów do służby w Agencji Bezpieczeństwa Wewnętrznego (t.j. DzU z 2014 r. poz. 61).

Dr Marcin Gołaszewski _____

Doktor nauk społecznych w dyscyplinie nauki prawne, biegły z zakresu badań poligraficznych przy Sądzie Okręgowym w Warszawie, prezes Zarządu Polskiego Towarzystwa Badań Poligraficznych.