

Tomasz Gólczewski

Instytut Biocybernetyki i Inżynierii Biomedycznej im. Macieja Nałęcz, PAN
Zakład Matematycznego Modelowania Procesów Fizjologicznych

Spirometryczne równania predykcyjne to rodzaj skamieliny — odpowiedź na komentarz T. Targowskiego

Spirometric prediction equations are a fossil
— reply to comments of T. Targowski

Pneumonol. Alergol. Pol. 2011; 80, 2: 189–192

Na wstępie pragnę podziękować za docenienie mojej „nowatorskiej fizjologicznie interpretowalnej konstrukcji równań”. Jako recenzent prac innych autorów widzę, że brak tego typu równań w tych pracach znacznie zmniejsza możliwość omawiania wyników (byłoby jednak rzeczą nieetyczną, gdybym jako recenzent wymuszał cytowanie własnych prac).

Krytyczne uwagi zawarte w komentarzu prof. Tomasza Targowskiego można podzielić na trzy grupy: metodologiczno-matematyczną, etyczną i dotyczącą technicznych aspektów badań pacjentów.

Co do kwestii metodologiczno-matematycznej (statystyka jest jedną dziedzin matematyki) dotyczącej rzekomej nadreprezentatywności osób w starszym wieku, muszę z żalem skonstatować, że jednak nie udało mi się w jasny sposób wytłumaczyć w ostatnich akapitach dyskusji [1] problemu prawidłowego rozkładu wieku w analizowanej próbie. Być może rycina 1 z bardzo obszernym opisem ułatwi nieco zrozumienie problemu.

W opisie ryciny 1 omówiłem ideę oraz przyczyny konstruowania równań predykcyjnych. Jak widać, równania te tworzone głównie po to, by w przeszłości lekarze mogli samodzielnie obliczać wartość należną, np. za pomocą ołówka, a potem kalkulatora. Stąd prosta, liniowa forma równań ECSC [2]. W dzisiejszych czasach należną wartość określa mikroprocesor w spirometrze, można więc proponować nawet bardzo złożone formy

równań, jak np. ta, którą opracowali Falaschetti i wsp. [3]. Problem w tym, że skoro dziś wartości należne określa mikroprocesor, sama idea równań jest przeżytkiem, równania są rodzajem atawizmu, pozostałością z dawnych lat. Obecnie chyba jedynym uzasadnieniem dla konstrukcji równań predykcyjnych może być ich wykorzystanie w zastosowaniach innych niż obliczanie wartości należnych, co jednak jest możliwe jedynie wtedy, gdy mają one formę interpretowalną, np. fizjologicznie, jak nasze równania [4].

Jeśli by jednak konstruować równania, interpretowalne czy nie, to — z punktu widzenia poprawności matematycznej — współczynniki członów określających wpływ wieku na wartość należną, tj. wartość średnią, powinny być określone za pomocą regresji wartości średnich dla poszczególnych roczników względem wieku. A więc byłaby jedna dana dla wieku 20 lat, jedna dana dla wieku 21 lat, ..., jedna dana dla wieku 90 lat, czyli każdy rocznik byłby „tak samo ważny” (co przy okazji byłoby zgodne z etyką lekarską). Oznacza to, że tworzenie równań na podstawie danych pacjentów powinno być dwustopniowe: najpierw powinno się obliczyć średnie dla wszystkich roczników, a dopiero potem przystąpić do regresji dającej matematyczny zapis zależności wartości średniej od wieku. W praktyce postępuje się inaczej: regresję wykonuje się bezpośrednio na „surowych” danych. Choć nie jest to całkiem poprawne z for-

Adres do korespondencji: dr hab. inż. Tomasz Gólczewski, ul. Trojdena 4, 02–109 Warszawa, tel.: 507 566 231, faks: (22) 658 28 72, e-mail: tgol@ibib.waw.pl

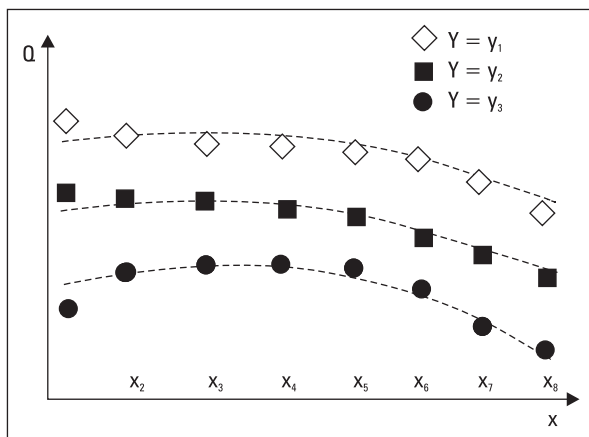
malnego punktu widzenia, to jednak również może prowadzić do tego samego poprawnego wyniku, jeśli w dalszym ciągu każdy rocznik był „tak samo ważny”, tj. gdyby liczba osób we wszystkich rozważanych rocznikach była taka sama. Dlatego podkreślam potrzebę w miarę jednorodnego rozkładu wieku w próbkę, mimo że rozkład w populacji ogólnej taki nie jest.

Wykorzystując metodę *reductio ad absurdum*, jeden z najpopularniejszych sposobów dowodzenia twierdzeń matematycznych, można w inny sposób zilustrować fakt, że rozkład wieku w próbce, na podstawie której określa się współczynniki równań predykcyjnych, nie może być taki sam jak w populacji ogólnej. Rzeczywiście, założmy wstępnie, że można skonstruować równanie, które prawidłowo określa wartość należną dla osób w wieku X , gdy takich osób w próbce było procentowo tyle samo co w populacji ogólnej. Liczba osób w wieku $X = 200$ lat w populacji ogólnej jest równa zero. Jednak z przyjętego założenia wynika, że można określić równanie predykcyjne wyliczające prawidłowo wartość należną dla 200-latków, bo w próbce nie mamy żadnej takiej osoby, czyli mamy ich tyle, ile jest w populacji ogólnej. Jest to raczej absurdalny wniosek, a więc założenie musi być fałszywe.

Aby indeksy spirometryczne dla osób młodych były określone równie dokładnie jak dla osób w starszym wieku, w analizowanej próbce liczby osób młodych i starych muszą być porównywalne. Jak można zobaczyć na rysunkach [1, 4], liczba i rozkład wiekowy osób młodych i starych są w przybliżeniu takie same i dlatego nie jest jasne, gdzie prof. Targowski znalazł dane świadczące o tym, że rzekomo 95% osób w analizowanej próbce jest w wieku powyżej 39 lat.

W odniesieniu do kwestii technicznych aspektów badań pacjentów można przyjąć poniższe wnioski.

Z komentarza prof. Targowskiego można by sądzić, że przez trzy lata wykonywał on badania pacjentów w sposób, o którym był lub powinien być przekonany, że jest sposobem niesolidnym. Wszak prof. Targowski powinien był i wtedy, i dziś wiedzieć, że to przede wszystkim właśnie diagnostyka konkretnych osób pod kątem identyfikacji choroby obturacyjnej wymaga dokładnego określenia wzrostu pacjenta; przecież ta sama wartość indeksu u osoby niskiej może być normą, a u osoby wysokiej świadczyć o chorobie. Dlatego nie sposób zrozumieć stwierdzenie pana profesora, że w trakcie badań w ramach „Nadzieja dla płuc” nie zaplanowano dokładnych pomiarów wzrostu i masy ciała, gdyż celem była „...identyfikacja chorych”.



Rycina 1. Jeśli fizjologiczna wielkość Q zależy od kilku parametrów X ($= x_1, x_2, \dots$), Y ($= y_1, y_2, \dots$), Z ($= z_1, z_2, \dots$), ..., to nie może istnieć jedna wartość średnia (należna) Q dla całej ludzkiej populacji, która mogłaby być wykorzystana w diagnostyce medycznej. Aby określić wartości Q , które byłyby użyteczne w medycynie, należałoby ludzką populację podzielić na subpopulacje określone przez kombinacje wartości parametrów X, Y, Z, \dots , tj. subpopulacje odpowiadające $[x_1, y_1, z_1, \dots]$, $[x_2, y_1, z_1, \dots]$, $[x_2, y_2, z_1, \dots]$, etc. Jeśli liczba możliwych wartości parametru Y jest znacznie mniejsza niż parametru X , średnie wartości Q mogą być przedstawione graficznie jak na tym rysunku. Lekarze mogliby wykorzystywać w diagnostyce taki diagram albo zbiór liczbowych danych zebranych w odpowiedniej tabeli. Jeśli jednak liczba możliwych subpopulacji jest ogromna, taka forma prezentacji danych byłaby niewygodna. Z tego powodu wybrane parametry traktuje się jako zmienne niezależne w pewnych równaniach opisujących zależność wartości średniej Q od wartości tych wybranych parametrów. Na przykład, linie kreskowane na rysunku reprezentują równania, w których parametr X pełni funkcję zmiennej niezależnej. Należy zauważyć, że te równania nie pozwalają obliczyć wartości średnich zbyt precyzyjnie (precyzja zależy od przyjętej formy równań; stąd taki nacisk na testowanie właściwej formy równań we wcześniejszych pracach [1, 4]).

Mając równania, musimy znać tylko kilka współczynników równań zamiast ogromnej bazy danych. Wynika stąd, że równania są mniej precyzyjne, ale za to bardziej wygodne, jeśli człowiek chce samodzielnie określić wartość średnią Q dla konkretnego pacjenta. Jednak dziś to nie lekarz, a mikroprocesor w spirometrze określa tę wartość. A dla niego nie ma różnicy, czy wykorzystuje nieprecyzyjne równania czy ogromną tablicę z precyzyjnymi danymi dla każdej subpopulacji. Tak więc równania predykcyjne mogą być dziś traktowane jako rodzaj skamieliny.

Jeśli Q jest indeksem spirometrycznym, np. FEV1, to X może oznaczać wiek (więc wartości x_1, x_2, x_3, \dots są wiekiem poszczególnych roczników), Y — rasę lub grupę etniczną (np. $y_1 = 1 =$ biali, $y_2 = 2 =$ Latynosi itd.), Z — płeć ($z_1 = 1 =$ mężczyźni, $z_2 = 2 =$ kobiety). Jest możliwe skonstruowanie jednego, ogólnego równania predykcyjnego dla całej ludzkości. Jednak takie równanie pozwalałoby na niezbyt precyzyjne określenie wartości należnej FEV1 dla konkretnej subpopulacji, tj. dla osób w tym samym wieku, o tej samej płci i rasie (oraz wzroście, który tu nie jest rozważany). Dwa równania, osobne dla z_1 i z_2 (tj. osobne dla mężczyzn i kobiet), byłyby lepsze, ale i one nie byłyby zbyt dokładne. Obecnie standardem jest zbiór równań opracowanych osobno dla każdej kombinacji wartości parametrów Y i Z , tj. dla $[y_1, z_1] =$ biali mężczyźni, $[y_1, z_2] =$ białe kobiety, $[y_2, z_1] =$ latynoscscy mężczyźni itd. Jednak to wciąż nie jest w pełni precyzyjne rozwiązanie. Rzeczywiście, współczynniki równania zależą od wartości FEV1 dla wszystkich wartości X (tj. wartości wieku). Dlatego, wartość należna FEV1 wyliczona dla danego wieku (np. dla osoby w starszym wieku) zależy od wartości FEV1 dla wszystkich pozostałych roczników, w tym np. osób młodych. A przecież wartość należna FEV1 dla osób w podeszłym wieku powinna zależeć jedynie od tego, jaką średnią wartością FEV1 cechują się osoby starsze.

W przypadku wszelkiego rodzaju analiz epidemiologicznych, w tym spirometrycznych równań predykcyjnych, byłoby dobrze, gdyby wszystkie niezbędne wielkości były określone (zmierzone) w miarę dokładnie. Dlatego szkoda, że wzrost nie był precyzyjnie mierzony. Ale wbrew stwierdzeniom prof. Targowskiego, to właśnie badanie konkretnych pacjentów, a nie określanie równań predykcyjnych, wymaga większej staranności (tu: starannego określenia wzrostu). W przypadku badań epidemiologicznych mniejsza precyzja wprowadzająca dodatkowy błąd przypadkowy nie ma zazwyczaj większego wpływu na tzw. momenty zwykłe — w naszym przypadku na wartość średnią indeksu spirometrycznego, czyli jego wartość należną. Taki błąd przypadkowy zanika bowiem podczas uśredniania i zwiększa jedynie przedział ufności danego współczynnika równania.

Fakt, że w ramach projektu „Nadzieja dla płuc” zbadano wielu palaczy oraz chorych lub podejrzewanych o chorobę, nie ma znaczenia, bowiem w analizach prowadzących do opracowania równań predykcyjnych uwzględniliśmy jedynie grupę osób zdrowych. Wszak, zgodnie z podanymi w artykule kryteriami wyłączenia [1, 4], odrzuciliśmy dane 76% mężczyzn i 69% kobiet (Falaschetti i wsp. [3], wykorzystując podobną, choć liczniejszą, bazę danych i stosując identyczne kryteria wyłączenia, odrzucili jeszcze większą liczbę osób: odpowiednio 82% i 79%).

Teoretycznie wiadomo, że wyniki spirometrii zależą od pory dnia i roku. Problem w tym, że w praktyce codziennej pacjentów bada się w każdej porze roku, a chyba w żadnych równaniach na wartości należne nie występuje współczynnik modyfikujący wartość należną w zależności od sezonu. Innymi słowy, wszystkie stosowane obecnie równania są równaniami ‘wielosezonowymi’. Dlatego, jeśli rzeczywiście — jak twierdzi prof. Targowski — badania w ramach „Nadzieja dla płuc” były wykonywane od maja do października, to nasze równania są lepsze, bardziej uniwersalne, bardziej ‘wielosezonowe’, niż myślałem (śp. dr. Lubiński podał [4], że badania były wykonywane do *September*, czyli do września, a nie do sierpnia).

Powyższe stanowisko jest ustosunkowaniem się inżyniera i matematyka do postawionych zarzutów dotyczących przydatności bazy danych projektu „Nadzieja dla płuc”. Korzystne byłoby też się zapoznać z opinią lekarzy kierujących tym projektem, tj. autora i kierownika projektu oraz koordynatora badań.

Opinia autora i kierownika projektu, czyli prof. Tadeusza Płusy, na temat przydatności wyników projektu wydaje się jednoznacznie pozytywna. Pan profesor zaproponował nam, pracownikom Instytutu Biocybernetyki i Inżynierii Biomedycznej PAN, byśmy włączyli się w opracowywanie

przez dr. Wojciecha Lubińskiego należnych spirometrycznych dla populacji polskiej na podstawie baz danych, które były w posiadaniu WIM, a które powstały przy różnych okazjach (tylko jedną z tych baz danych stanowiły wyniki projektu „Nadzieja dla płuc”). Propozycja padła na spotkaniu skutkującym oficjalną umową o współpracy naukowej pomiędzy WIM i IBIB PAN. Naszą stroną reprezentował dyrektor ds. naukowych IBIB PAN, prof. Marek Darowski, któremu towarzyszyli autor oraz mgr inż. Krzysztof J. Pałko. Jeśli więc prof. Płusa, wówczas dyrektor ds. naukowych WIM, a jednocześnie autor i kierownik projektu złożył tę propozycję, musiał być przekonany o przydatności baz danych z medycznego, metodologicznego punktu widzenia.

Należy jednak dodać, że przeprowadzając analizy matematyczne udostępnionych mi baz danych, stwierdziłem statystycznie istotne różnice między bazami z XX i XXI w., prawdopodobnie związane ze zmianami w wytycznych dotyczących przeprowadzania badania spirometrycznego (ciekawe studia, oparte na analizie ww. różnic między bazami, dotyczące ewentualnych przyczyn ‘historycznych’ zmian wartości należnych, a zwłaszcza zmiany podejścia do zależności FEV1/FVC od wieku, zostały przerwane przez tragiczną śmierć dr. Lubińskiego). Z powodu tych różnic ostatecznie wykorzystana została jedynie najbardziej aktualna baza danych, tj. baza danych projektu „Nadzieja dla płuc”, co jednak nie oznacza, że opracowanie równań było jakkolwiek kontynuacją tego projektu.

Oczywiście opinia koordynatora badań, dr. Wojciecha Lubińskiego, również musiała być pozytywna, skoro — zachęcany przez prof. Tadeusza Płusę, swojego opiekuna naukowego — podjął się on opracowania równań predykcyjnych dla populacji polskiej. Skoro obaj lekarze-naukowcy kierujący projektem „Nadzieja dla płuc” byli przekonani o poprawności wykonywania badań, to nie miałem podstaw, by w to wątpić, i dlatego w swojej pracy w *Pneumonologii i Alergologii Polskiej* [1] podał (a nie: potwierdziłem), że badania były wykonane zgodnie z obecnymi wytycznymi (bazy danych z XX w., które odrzuciłem, nie były używane w sposób zgodny z obecnymi wytycznymi).

Co do kwestii etycznej, tj. rzekomego braku podziękowań dla pana prof. Tadeusza Płusy w pracy opublikowanej w *Journal of Applied Physiology* [4], to z przykrością muszę stwierdzić, że pan prof. Targowski widocznie nie zapoznał się dokładnie z tą pracą, bowiem w części ACKNOWLEDGMENTS są zamieszczone podziękowania dla B. Stankiewicz oraz M.D. Tadeusza Płusy (stopnie i tytuły naukowe obu osób usunęła redakcja czasopisma).

W przypadku artykułu opublikowanego w *Pneumonologii i Alergologii Polskiej* [1] opierałem się na trzech wcześniej opublikowanych pracach [2–4]. Nie spotkałem się jeszcze z przypadkiem, by ktoś dziękował autorom zacytowanych prac. Gdybym jednak umieścił podziękowania osobom związanym w jakiś sposób z jedną z tych prac, musiałbym również zamieścić podziękowania dla osób związanych z dwiema pozostałymi pracami, tj. dla Falaschetti, Quanjer oraz ich współpracowników. A to już byłoby chyba znaczną przesadą.

Piśmiennictwo

1. Gólczewski T. Spirometria — porównanie polskich wartości należnych Lubińskiego z należnymi ECSC/ERS i Falaschetti. *Pneumonol. Alergol. Pol.* 2012; 80, 1: 29–40.
2. Quanjer P.H., Tammeling G.J., Cotes J.E., Pedersen O.F., Peslin R., Yernault J.C. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur. Respir. J. Suppl.* 1993; 16: 5–40.
3. Falaschetti E., Laiho J., Primates P., Purdon S. Prediction equations for normal and low lung function from the Health Survey for England. *Eur. Respir. J.* 2004; 23: 456–463.
4. Lubiński W., Gólczewski T. Physiologically interpretable prediction equations for spirometric indices. *J. Appl. Physiol.* 2010; 108: 1440–1446.