



dr SZYMON ORDYSIŃSKI (ORCID 0000-0002-5907-6582)
 Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy
 Kontakt: szord@ciop.pl
 DOI: 10.54215/BP.2023.10.21.Ordysinski

Szacowanie prawdopodobieństwa wystąpienia i ciężkości wypadku przy pracy na podstawie analizy danych statystycznych

Fot. Attapornfoto/Bigstockphoto



Artykuł przedstawia sposób wykorzystania metod analizy danych statystycznych, rejestrowanych w ogólnokrajowych bazach, dotyczących wypadków przy pracy i osób pracujących. Wyniki takiej analizy mogą służyć do określania prawdopodobieństwa wystąpienia i ciężkości wypadku przy pracy w poszczególnych grupach pracowników, zdefiniowanych na podstawie cech tych pracowników i okoliczności wykonywanej przez nich pracy. Przeprowadzona analiza danych statystycznych pochodzących z dwóch różnych baz, w której posłużono się specjalnie opracowaną metodą integracji danych i identyfikacji reprezentatywnych grup pracowników (polegającą na stopniowym ograniczaniu reguły podziału i wykorzystaniu współczynnika zmienności *bootstrap*), pozwoliła na obliczenie wskaźnika prawdopodobieństwa wypadku przy pracy. Natomiast połączenie kilku jednowymiarowych i wielowymiarowych metod analizy danych umożliwiło zidentyfikowanie grup poszkodowanych w wypadkach przy pracy, które są wystarczająco liczne i silnie zróżnicowane pod względem ciężkości wypadków, co zapewnia wysoką trafność przewidywanej na ich podstawie ciężkości wypadków.

Słowa kluczowe: wypadki przy pracy, dane statystyczne, prawdopodobieństwo wypadku, ciężkość wypadku, analiza danych

Estimation of the probability and the severity of an accident at work based on statistical data analysis

The article presents a method of analysing statistical data registered in national databases concerning accidents at work and persons employed. The results of such analysis can be used to determine the probability and severity of accidents at work in particular groups of employees, defined on the basis of the characteristics of these employees and the circumstances of their work. The analysis of statistical data from two different databases, in which a specially developed method of data integration and identification of representative groups of employees was used (consisting in the gradual reduction of the division rule and the use of the *bootstrap* coefficient of variation), allows for the calculation of the probability index of occupational accidents. In addition, the combination of several one-dimensional and multidimensional analysis methods made it possible to identify groups of injured in accidents at work that are sufficiently numerous and strongly diversified in terms of the severity of accidents, which ensures high accuracy of the severity of accidents predicted on their basis.

Keywords: accidents at work, statistical data, accident probability, accident severity, data analysis

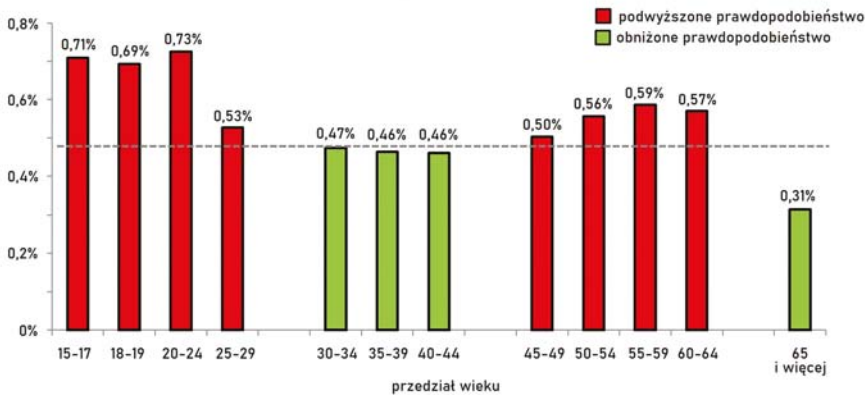
Wstęp

Dane statystyczne dotyczące wypadków przy pracy, rejestrowane w ogólnokrajowych bazach danych wszystkich krajów członkowskich Unii Europejskiej, mogą być istotnym źródłem rzetelnej i potwierdzonej empirycznie wiedzy na temat prawdopodobieństwa powstawania i ciężkości wypadków przy pracy. Obecnie jednak tego typu dane – np. udostępniane przez Eurostat i GUS oraz przytaczane w krajowych publikacjach – są prezentowane w formie, która – ze względu na zbyt uogólnione opisy grup pracowników czy nie dość szczegółowo opisane stanowiska pracy – bardzo utrudnia ich wykorzystanie do formułowania wniosków na temat prawdopodobieństwa wystąpienia i ciężkości wypadków w określonej grupie pracowników. Zastosowanie nowego podejścia do analizy tych danych oraz prezentowania wyników, z uwzględnieniem poszczególnych grup pracowników, szczególnie wyróżnionych na podstawie cech tych pracowników i okoliczności wykonywanej przez nich pracy, może się przyczynić do bardziej powszechnego wykorzystywania tego typu danych podczas przeprowadzania oceny ryzyka zawodowego i planowania działań prewencyjnych na poziomie przedsiębiorstwa i kraju.

Celem artykułu jest prezentacja wybranej metody analizy danych statystycznych dotyczących wypadków przy pracy i osób pracujących, której wyniki mogą stanowić wiarygodne źródło potwierdzonej empirycznie wiedzy o prawdopodobieństwie występowania i ciężkości wypadków w określonych grupach pracowników. Wiedza ta z powodzeniem może być wykorzystywana w różnego rodzaju działaniach prewencyjnych.

Źródła danych o wypadkach przy pracy i pracujących w gospodarce narodowej

Analiza ciężkości wypadków przy pracy powstających w poszczególnych grupach pracowników może być wykonana na podstawie bazy danych dotyczących wypadków przy pracy. Przeprowadzenie analizy prawdopodobieństwa powstania wypadku wymaga jednak dysponowania dodatkową informacją o liczbie pracujących w danej grupie.



Rys. 1. Odsetek osób poszkodowanych w wypadkach przy pracy w poszczególnych grupach wiekowych – na podstawie danych GUS, zarejestrowanych w latach 2017-2019

Fig. 1. Percentage of people injured in accidents at work in particular age groups – based on data from the Central Statistical Office, registered in 2017-2019

Najbardziej odpowiednie źródło informacji o wypadkach przy pracy stanowi baza danych rejestrowanych przez GUS w ramach „Statystycznej karty wypadku Z-KW” [1], ponieważ uwzględnia się w niej szeroki zakres informacji dotyczących okoliczności wypadku, a co najważniejsze – wszystkie rodzaje wypadków przy pracy (niezależnie od powodowanej przez nie długości niezdolności do pracy). Z kolei za podstawowe źródło danych o pracujących w gospodarce narodowej należy uznać Badanie Aktywności Ekonomicznej Ludności (BAEL) [2] – głównie z uwagi na obszerność rejestrowanych informacji i dużą próbę badawczą, która zapewnia reprezentatywność wyników dla poszczególnych grup zawodowych. Dane dotyczące pracujących oraz dane dotyczące poszkodowanych w wypadkach przy pracy są zatem rejestrowane w różnych bazach danych, w ramach osobnych badań. Przeprowadzenie analiz prawdopodobieństwa powstania wypadku wymaga więc zintegrowania bazy danych o poszkodowanych w wypadkach przy pracy z bazą zawierającą informacje o pracujących.

Analizy prawdopodobieństwa i ciężkości wypadku przeprowadzono na podstawie danych zarejestrowanych w bazach ZK-W i BAEL w latach 2017-2019. Przyjęcie w analizach dłuższego, trzyletniego okresu pozwala zredukować wpływ losowo występujących zmian, które nie wynikają z ogólnych tendencji.

Szacowanie prawdopodobieństwa wypadku przy pracy

Na potrzeby oszacowania prawdopodobieństwa wypadku przy pracy konieczne było zintegrowanie danych pochodzących z dwóch różnych baz. Wymagało to przede wszystkim pogrupowania obserwacji zawartych w tych bazach w taki sposób, by odnosiły się do tych samych populacji pracujących i poszkodowanych w wypadkach przy pracy. Spośród 18 zmiennych opisujących okoliczności wypadku przy pracy tylko siedem znajduje swoje odpowiedniki wśród niemal 100 zmiennych rejestrowanych w ramach badania BAEL.

Integracja baz danych jest zatem możliwa tylko w zakresie zmiennych opisujących w kolejności: rodzaj działalności PKD (według klasyfikacji), zawód wykonywany (według klasyfikacji), wielkość przedsiębiorstwa, status zatrudnienia, płeć, wiek i staż pracy pracownika.

Identyfikacja grup osób pracujących na podstawie pełnego zakresu kodów oraz wszystkich zmiennych zidentyfikowanych w obu bazach prowadziłaby do zbyt szczegółowego podziału obserwacji – analizowane grupy nie miałyby wystarczająco dużej liczebności, by móc formułować rzetelne wnioski. Konieczne było więc ograniczenie kryteriów podziału pracujących na grupy w sposób niepowodujący utraty kluczowych informacji. Z tego względu stopniowo ograniczono reguły podziału, rozpoczynając od najmniej istotnych informacji, a więc tych, które mają najmniejszy wpływ na różnicowanie prawdopodobieństwa w poszczególnych grupach pracowników.

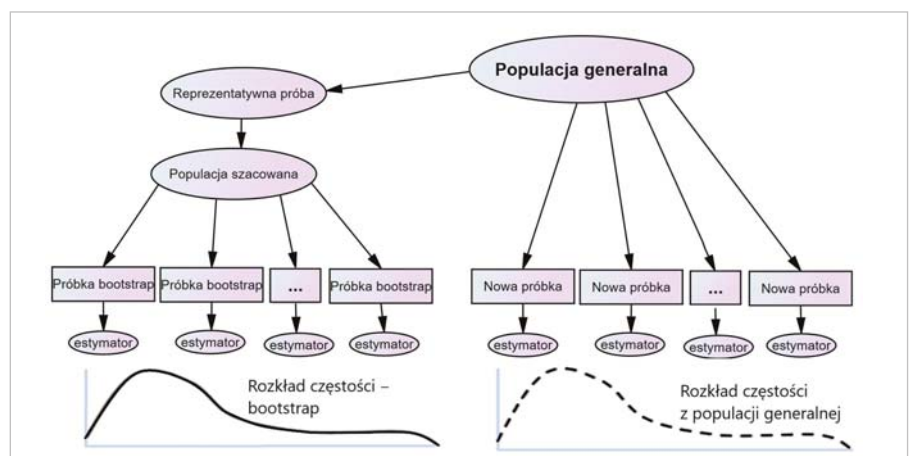
W tym celu przede wszystkim przekształcono zmienną „wiek pracowników”. Analiza prawdopodobieństwa powstania wypadku w poszczególnych grupach wiekowych wykazała, że zmienność prawdopodobieństwa można w wystarczający sposób wyrazić, dzieląc poszkodowanych jedynie na dwie grupy wiekowe – o zwiększonym i obniżonym

prawdopodobieństwie wypadku. Pierwsza obejmuje pracowników do 29. roku życia oraz pomiędzy 45. a 64. rokiem życia (są to osoby, w przypadku których istnieje zwiększone ryzyko wypadku przy pracy), natomiast pozostali pracownicy zostali zakwalifikowani do drugiej grupy (rys. 1).

Uznano też, że można zrezygnować ze zmiennej „staż pracy”, która jest najmniej istotna w wyjaśnieniu zmienności prawdopodobieństwa wypadku. Podobnie wielkość przedsiębiorstwa nie jest najlepszym predyktorem prawdopodobieństwa powstania wypadku. Ponadto ograniczono zmienną „zawód wykonywany przez pracownika”, wykorzystując w tym celu bardziej szczegółową informację o zawodzie wyuczonym oraz specjalnie opracowany algorytm, który informację o zawodzie wykonywanym uszczęławia tylko w określonych sytuacjach – gdy występuje wysokie prawdopodobieństwo zgodności zawodu wykonywanego z zawodem wyuczonym. Umożliwia to identyfikację grup zawodów zróżnicowanych ze względu na specyficzny zakres obowiązków, proces pracy i zagrożenia wypadkowe. Przykładem takiej grupy zawodowej są specjaliści do spraw zdrowia, wśród których znajdują się lekarze i pielęgniarki różnych specjalizacji, ale również ratownicy medyczni, rehabilitanci, farmaceuci oraz lekarze weterynarii, których cechują nieco inny proces pracy oraz odmienne ryzyko zawodowe. Z tego powodu w modelu przewidywania prawdopodobieństwa wypadku te grupy powinny być identyfikowane oddzielnie.

Dalsze ograniczanie liczby zastosowanych zmiennych lub wartości ich klasyfikacji nie jest wskazane z uwagi na zbyt dużą utratę informacji i znajduje uzasadnienie tylko wobec pojedynczych, dokładnie zidentyfikowanych grup pracowników, które nie spełniają kryterium reprezentatywności. W celu identyfikacji niereprezentatywnych grup zastosowano współczynnik zmienności *bootstrap* (oparty na metodzie *bootstrap*), którego wysokie wyniki umożliwiają identyfikację grup pracowników o niskiej reprezentatywności w próbce [3-4].

Metoda *bootstrap* służy do poprawy oszacowania rozkładu analizowanych estymatorów w populacji generalnej poprzez określenie spodziewanego



Rys. 2. Schemat postępowania w metodzie *bootstrap*
Fig. 2. The bootstrap method scheme

rozproszenia – na podstawie obliczenia wariancji *bootstrap* w odniesieniu do każdego pomiaru. Zastawanie odbywa się poprzez ponowne wykorzystanie obserwowanej próby badania, poprzez pobranie nowych, losowo wygenerowanych próbek, *in silico* (na komputerze), a następnie akumulację wszystkich pozyskanych wartości i utworzenie w ten sposób nowego rozkładu, który w teorii odzwierciedla rozkład częstości występowania zidentyfikowanych grup pracowników w populacji generalnej. Poszczególne etapy realizacji metody *bootstrap* zaprezentowano na rys. 2.

Obliczenie wariancji *bootstrap* dla każdej grupy pracowników umożliwia następnie określenie współczynnika zmienności CV, zgodnie ze wzorem [5]:

$$\bigwedge_{g \in n} CV(\hat{t}) = 100\% \cdot \frac{\sqrt{\hat{V}(\hat{t})}}{\hat{t}_s}$$

gdzie:

$$\hat{V}(\hat{t}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{t}_b^* - \hat{t}_s)^2,$$

$$\hat{t}_s = \frac{1}{B} \sum_{b=1}^B \hat{t}_b^*.$$

\hat{t}_b – wartość kolejnych wygenerowanych estymatorów (liczba pracujących w poszczególnych grupach),

n – wielkość próby,

$B = 1000$ – liczba próbek *bootstrap*.

Bazując na wynikach współczynnika zmienności, zidentyfikowano grupy o niskiej reprezentatywności w próbce i połączono je w większe skupiska dzięki przekształceniom, które stopniowo ograniczają szczegółowość kryteriów identyfikacji – od najmniej istotnej informacji, w następującej kolejności: płeć, wiek, zawód – przy czym dotyczy to tylko niereprezentatywnych grup. W efekcie zastosowania tej metody z niemal 5 tys. początkowo zidentyfikowanych grup pracowników po przekształceniach otrzymano ok. 2,5 tys. grup reprezentatywnych, które są opisane przez siedem zmiennych o różnej szczegółowości informacji.

Obliczenie współczynnika zmienności umożliwiło również obliczenie dolnej i górnej granicy przedziału ufności liczby pracujących dla każdej zidentyfikowanej grupy pracowników [6]:

$$\hat{t} \cdot \left(1 - 1,96 \frac{CV(\hat{t})}{100}\right) \leq t \leq \hat{t} \cdot \left(1 + 1,96 \frac{CV(\hat{t})}{100}\right)$$

Przedziały ufności umożliwiają wskazanie, że realna liczba pracowników w populacji generalnej będzie się mieściła w granicach przedziału ufności, z określonym (95%) prawdopodobieństwem. Przy znanej liczbie wypadków oszacowanie to przekłada się również na prawdopodobieństwo wypadku. Informacja ta jest bardzo pomocna w interpretacji wyników – nie tylko dlatego, że zwiększa rzetelność szacowania prawdopodobieństwa, lecz także dlatego, że wskazuje, jak duża jest precyzja wyników.

Na podstawie tak przekształconej bazy możliwe było – po zastosowaniu odpowiednich wag (mnożników uogólniających i wskaźników korygujących) – określenie rzetelnej liczby pracujących w grupach sklasyfikowanych według rodzaju działalności, wieku, płci oraz zawodu (opisanego odpowiednio za pomocą cztero-, trzy- lub dwuliczbowego kodu). Informacje te następnie zestawiono z liczbą poszkodowanych w wypadkach przy pracy w zidentyfikowanych grupach i na tej podstawie obliczono zakres prawdopodobieństwa powstania wypadku, wynikający z obliczonych przedziałów ufności dla liczby pracowników.

Szacowanie ciężkości wypadku przy pracy

Do opracowania modelu statystycznego, który na podstawie cech poszkodowanych w wypadkach przy pracy oraz innych okoliczności pracy wykonywanej przez te osoby umożliwiłby wiarygodne przewidywanie ciężkości wypadku przy pracy, wykorzystano połączenie metod jedno- i wielowymiarowych.

Metody jednowymiarowe zastosowano przede wszystkim do oceny siły związku ciężkości wypadków przy pracy (mierzonej liczbą dni niezdolności do pracy) z kolejnymi cechami pracowników i okolicznościami wykonywanej przez nich pracy. Do oceny tej skuteczności zastosowano miary wielkości efektu – d Cohena dla homogenicznych grup [7]. W przeprowadzonej analizie zastosowano różne miary, dostosowane do poziomu pomiaru analizowanych zmiennych, do analizy predyktorów: dichotomicznych (np. płeć) d Cohena, ilościowych (np. wiek) r-kwadrat, nominalnych o wielu wartościach (np. zawód wykonywany) eta-kwadrat. W celu porównania wyników odmiennych miar wielkości efektu konieczne było przeliczenie ich wartości na wspólną miarę d Cohena. Pozwoliło to uzyskać jednolitą interpretację siły zidentyfikowanych związków [8-10].

W celu ograniczenia liczby wartości poszczególnych zmiennych zastosowano metodę drzew CHAID (*Chi-squared Automatic Interaction Detection*) [11-12]. Do dalszego ograniczania kryteriów podziału, zarówno wobec liczby zmiennych, jak i ich wartości, zastosowano metodę drzew klasyfikacyjnych CRT (*Classification and Regression Trees*)

[11-12]. Umożliwia ona wielowymiarowe sklasyfikowanie analizowanych obserwacji przez liczne następujące po sobie podziały badanej zbiorowości, z wykorzystaniem przy każdym kolejnym podziale pojedynczej zmiennej, której wartości najsilniej różnicują (ze względu na wartość zmiennej zależnej) fragment zbiorowości aktualnie poddawany podziałowi. Ostatecznie zidentyfikowane grupy zostają opisane przez wybrane wartości kilku zmiennych niezależnych. Ponieważ przy każdym kolejnym podziale następował nowy dobór zmiennej i jej wartości – w taki sposób, by jak najskuteczniej podzielił badaną zbiorowość ze względu na zróżnicowanie wartości zmiennej zależnej – dlatego zidentyfikowane grupy są opisane przez określony zbiór wybranych wartości poszczególnych zmiennych.

Etapy analiz

W ramach przeprowadzonych analiz przede wszystkim rozpoznano jednowymiarowy związek okoliczności pracy, wykonywanej przez poszkodowanego w chwili wypadku, z ciężkością wypadków przy pracy. Do analizy siły tego związku zastosowano różne miary wielkości efektu, dostosowane do poziomu pomiaru analizowanych zmiennych, których wartość następnie przeliczono na wspólną miarę d Cohena w celu zapewnienia jednolitej interpretacji siły zidentyfikowanych związków (zgodnie z: [8-10]). W toku tych analiz określono skuteczność przewidywania niezdolności do pracy na podstawie cech środowiska pracy i poszkodowanego pracownika oraz wykonywanej przez niego czynności. W ten sposób zidentyfikowano sześć zmiennych, które uzyskały oczekiwaną wartość miary wielkości d Cohena – na podstawie ich wartości można skutecznie przewidywać ciężkość wypadku przy pracy (zob. tabelę). W przypadku zmiennych o mniejszej wartości miary efektu wpływu należy uznać, że określone na ich podstawie grupy poszkodowanych nie różnią się pod względem długości absencji powypadkowej, co oznacza, że wartości tych zmiennych nie pozwalają na skuteczne przewidywanie okresu niezdolności do pracy spowodowanej wypadkiem przy pracy (na podstawie: [8-10]).

Dalsze analizy wykazały, że gdy do podziału poszkodowanych na grupy zastosuje się wartości wszystkich sześciu zmiennych zidentyfikowanych jako skuteczne predyktory ciężkości wypadku, nie

Tabela. Zmienne opisujące poszkodowanego i okoliczności wykonywanej przez niego pracy, które uzyskały najwyższy wynik miary wielkości efektu d Cohena

Table. Variables describing the injured person and the circumstances of their work that obtained the highest score on the Cohen d effect size measure

Predyktor ciężkości wypadku przy pracy	Wartość d Cohena
Czynnik materialny czynności	0,4825
PKD	0,4442
Zawód wykonywany	0,4291
Miejsce powstania wypadku	0,3577
Proces pracy	0,3203
Wiek poszkodowanego	0,3201

przynosi to pożądanego efektu, ponieważ liczba zidentyfikowanych w ten sposób grup poszkodowanych w wypadkach przy pracy staje się zbyt duża, a liczebność poszkodowanych wewnątrz grup – zbyt mała (w zdecydowanej większości powstają grupy jednoosobowe), by móc formułować rzetelne i trafne wnioski na temat ciężkości powstawania wypadków przy pracy. Z tego powodu postanowiono wprowadzić dalsze ograniczenia, już nie tylko w liczbie zmiennych, lecz także w zakresie wartości branych pod uwagę przy identyfikacji grup.

W celu ograniczenia liczby wartości poszczególnych zmiennych postanowiono przeprowadzić analizę drzew klasyfikacyjnych CHAID. Dzięki temu rozproszone grupy, wyznaczone przez nominalne wartości predyktorów, połączono w liczniejsze skupiska, bardziej jednorodne ze względu na długość absencji powypadkowej. W wyniku analizy powstały nowe zmienne, o znacznie ograniczonej liczbie wartości, jednak bez istotnej utraty możliwości wyjaśniania zmiennej zależnej (związek poszczególnych zmiennych z niezdolnością do pracy, wyrażony wartością miary eta kwadrat, pozostał na podobnym poziomie). Najbardziej udało się zredukować liczby wartości zmiennych zawodów wykonywanych (z ponad 2000 wartości do 24) oraz rodzaj działalności gospodarczej (z niemal 600 do 22).

Pomimo wprowadzenia przedstawionych ograniczeń dotyczących zmiennych użytych do identyfikacji grup pracowników liczba tych grup wciąż była zbyt duża, a ich liczebność w wielu przypadkach – zbyt mała. W kolejnej analizie zastosowano więc metodę drzew CRT, aby zidentyfikować grupy poszkodowanych w wypadkach przy pracy, które będą silnie zróżnicowane pod względem długości absencji powypadkowej. Opracowany algorytm podziału pozwolił wyodrębnić 264 grupy poszkodowanych w wypadkach, określone przez ograniczony zbiór wartości sześciu zidentyfikowanych i przekształconych zmiennych opisujących poszkodowanego i okoliczności wykonywanej przez niego pracy. Zidentyfikowane w ten sposób grupy są bardzo silnie zróżnicowane pod względem ciężkości wypadków, co potwierdza wysoki wynik miary wielkości efektu eta kwadrat = 0,101 (co odpowiada: d Cohena = 0,67) – najwyższy ze wszystkich analizowanych zmiennych – świadczący o silnym związku zidentyfikowanej w ten sposób klasyfikacji z liczbą dni niezdolności do pracy. Jednocześnie taki podział zapewnia wysoką liczebność wewnątrzgrupową i tym samym dużą wiarygodność wyników.

Narzędzie komputerowe umożliwiające przewidywanie prawdopodobieństwa i ciężkości wypadku przy pracy

Przeprowadzone analizy wykazały, że na podstawie dostępnych danych statystycznych możliwe jest przewidywanie prawdopodobieństwa i ciężkości wypadków przy pracy w grupach pracowników. Tego typu informacje mogą być przydatne podczas oceny ryzyka zawodowego, jednak ze względu na znaczne skomplikowanie interpretacja wyników

przez potencjalnych odbiorców jest utrudniona. Z tego powodu postanowiono opracować narzędzie komputerowe, umożliwiające użytkownikom odnajdowanie potrzebnych informacji.

Podstawą do stworzenia narzędzia komputerowego były opracowane modele przewidywania prawdopodobieństwa i ciężkości wypadku przy pracy – wyniki uzyskane za ich pomocą zostały zapisane w formie baz danych będących częścią tego narzędzia. Narzędzie zostało opracowane w języku programowania VBA for Excel, z użyciem makr i formularzy, z warstwą graficzną i odpowiednio opracowanym kodem źródłowym. Ostatecznie narzędzie składa się z pięciu plików, tj. głównego pliku startowego (typu Excel; *.xslm) i czterech zewnętrznych plików Excel (*.xslm), zawierających poszczególne bazy danych.

Po wprowadzeniu przez użytkowników danych o pracowniku i okolicznościach wykonywanej przez niego pracy program przeszukuje bazy danych i podaje przewidywane prawdopodobieństwo wypadku przy pracy i przewidywaną liczbę dni niezdolności do pracy, wraz z przedziałami ufności, pięciostopniową interpretacją wyników oraz interpretacją opisową. Interpretacja wyniku oparta jest na wartości centylowej danej grupy pracowników. Narzędzie podaje również najważniejsze elementy przebiegu najczęściej występujących wypadków przy pracy w zidentyfikowanej grupie.

Dla ciężkości wypadków opracowano następującą interpretację przedziałów:

1. Brak ciężkich uszkodzeń ciała, brak lub bardzo krótkotrwała niezdolność do pracy (od 0 do 21 dni),

2. Brak ciężkich uszkodzeń ciała, krótkotrwała niezdolność do pracy, niewielkie koszty wypadku (od 22 do 31 dni),
3. Mogą pojawić się uszkodzenia ciała skutkujące średnią niezdolnością do pracy. Koszty wypadku mogą być odczuwalne (od 32 do 41 dni),
4. Uszkodzenia ciała skutkujące relatywnie długą niezdolnością do pracy. Zdolność do pracy przywrócona po leczeniu. Wyższe koszty wypadku (od 42 do 49 dni),
5. Ciężkie uszkodzenia ciała skutkujące długą lub trwałą niezdolnością do pracy. Wysokie koszty wypadku (od 50 do 182 dni).

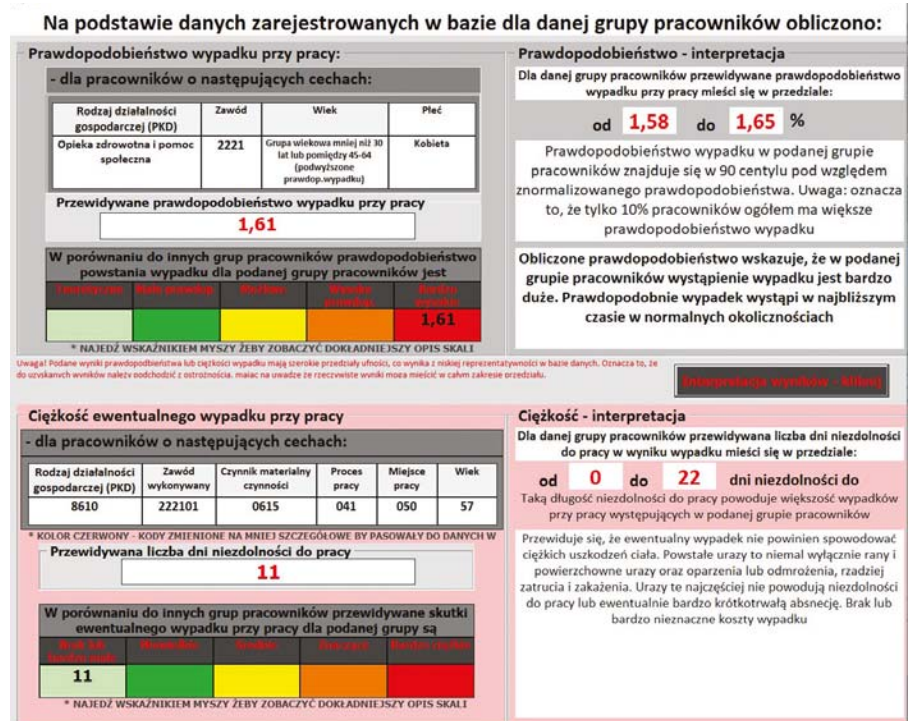
Przykładowe wyniki otrzymane na podstawie opracowanych modeli

Po wprowadzeniu do programu odpowiednich parametrów dotyczących pracownika i jego środowiska pracy za pomocą opracowanego narzędzia można otrzymać wyniki dotyczące prawdopodobieństwa i ciężkości wypadku przy pracy. Oto przykłady ilustrujące działanie tego narzędzia.

Przykład 1. Pielęgniarka

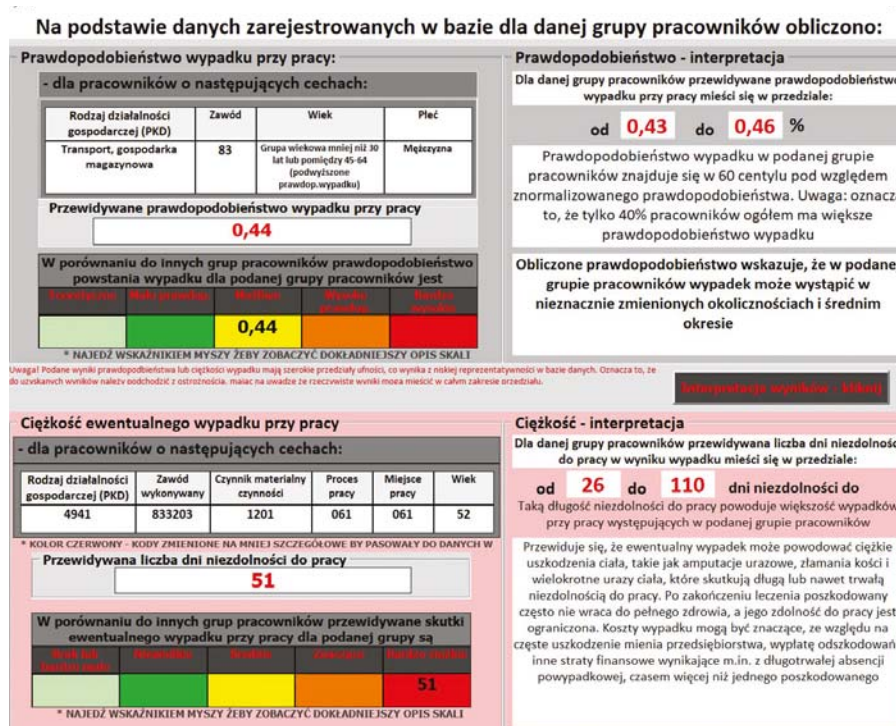
Wprowadzone parametry – pracownik i środowisko pracy:

- zawód: pielęgniarka,
- PKD: działalność szpitalna,
- miejsce pracy: placówka ochrony zdrowia,
- proces pracy: usługi, opieka, pomoc dla społeczeństwa,
- czynnik materialny czynności: narzędzia ręczne bez napędu – chirurgiczne i inne medyczne – nietnące, do prac medycznych,



Rys. 3. Wyniki prawdopodobieństwa i ciężkości wypadku, podane przez opracowane narzędzie komputerowe, dla przykładowego pracownika wykonującego zawód pielęgniarka

Fig. 3. The results of the probability and severity of an accident, given by the developed computer tool, for an example of a worker in the nursing profession



Rys. 4. Wyniki prawdopodobieństwa i ciężkości, podane przez opracowane narzędzie komputerowe, dla przykładowego pracownika wykonującego zawód kierowcy samochodu ciężarowego
Fig. 4. The results of the probability and severity of an accident, given by the developed computer tool, for an example of a worker in the occupation of truck driver

- wiek: 57 lat,
 - płeć: kobieta.
- Otrzymane wyniki (rys. 3):
- przewidywane prawdopodobieństwo powstania wypadku: 1,61% (poziom ufności = 90%: 1,58%-1,65%), interpretacja: bardzo wysokie,
 - przewidywana ciężkość ewentualnego wypadku: 11 dni (poziom ufności = 50%: 0-21 dni), interpretacja przewidywanych skutków: brak lub bardzo małe (brak ciężkich uszkodzeń ciała, brak niezdolności do pracy lub bardzo krótkotrwała niezdolność do pracy).

Przykład 2. Kierowca samochodu ciężarowego

- Wprowadzone parametry – pracownik i środowisko pracy:
- zawód: kierowca samochodu ciężarowego,
 - PKD: transport drogowy towarów,
 - miejsce pracy: teren komunikacji publicznej (np. drogi, parkingi, poczekalnie),
 - proces pracy: przemieszczanie się, w tym środkami transportu,
 - czynnik materialny czynności: ciężkie pojazdy transportowe,
 - wiek: 52 lata,
 - płeć: mężczyzna.
- Otrzymane wyniki (rys. 4):
- Przewidywane prawdopodobieństwo wypadku: 0,44% (poziom ufności = 90%: 0,43%-0,46%), interpretacja: możliwe,
 - Przewidywana ciężkość ewentualnego wypadku: 51 dni (poziom ufności = 50%: 26-110 dni), interpretacja przewidywanych skut-

ków: bardzo ciężkie (ciężkie uszkodzenia ciała skutkujące długą lub trwałą niezdolnością do pracy, wysokie koszty wypadku).

Podsumowanie

Przedstawione analizy wykazały, że na podstawie danych statystycznych rejestrowanych w ogólnokrajowych bazach można w sposób rzetelny określać prawdopodobieństwo powstawania i ciężkości wypadków przy pracy w poszczególnych grupach pracowników, określonych na podstawie cech tych pracowników i okoliczności wykonywanej przez nich pracy. Wyniki tych analiz mogą stanowić źródło potwierdzonej empirycznie wiedzy o wypadkowości w określonych grupach pracowników, przydatnej podczas przeprowadzania oceny ryzyka zawodowego oraz w innych działaniach prewencyjnych.

Obliczenie prawdopodobieństwa wypadku przy pracy w odniesieniu do reprezentatywnych grup pracowników wymagało przeprowadzenia integracji danych, ponieważ dane o wypadkach przy pracy i osobach pracujących pochodzą z dwóch różnych baz. W tym celu posłużono się specjalnie opracowaną metodą identyfikacji grup pracowników (polegającą na stopniowym ograniczaniu reguły podziału i wykorzystaniu wskaźnika zmienności *bootstrap*), która pozwoliła na zidentyfikowanie reprezentatywnych grup pracowników i obliczenie wiarygodnego wskaźnika prawdopodobieństwa powstawania wypadku przy pracy.

Przewidywanie ciężkości ewentualnego wypadku przy pracy w grupach pracowników określonych na podstawie cech tych pracowników i okoliczności wykonywanej przez nich pracy

wiązało się z koniecznością połączenia kilku jedno- i wielowymiarowych metod analizy. Pozwoliło to na zidentyfikowanie grup poszkodowanych w wypadkach przy pracy (określonych przez wybrane wartości sześciu zmiennych opisujących okoliczności pracy), które są wystarczająco liczne i silnie zróżnicowane pod względem ciężkości wypadków. Wysoka liczebność grup zapewnia wysoką wiarygodność i trafność przewidywanej na ich podstawie ciężkości wypadków.

BIBLIOGRAFIA

- [1] Zeszyt metodologiczny. Wypadki przy pracy. Gdańsk: GUS, 2021.
- [2] Zeszyt metodologiczny. Badanie Aktywności Ekonomicznej Ludności. Warszawa: GUS, 2018.
- [3] EFRON B. 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association. 1983, 78 (382): 316-331.
- [4] EFRON B., TIBSHIRANI R. Improvements on cross-validation: the 632+ *bootstrap* method. Journal of the American Statistical Association. 1997, 92(438): 548-560.
- [5] WOLTER K.M. Introduction to Variance Estimation. New York: Springer Verlag, 1985, 2007.
- [6] SÄRNDAL C.E., SWENSSON B., WRETMAN J. Model Assisted Survey Sampling. Springer New York, 1992.
- [7] COHEN J. Statistical power analysis for the behavioral sciences. Wyd. 2. Nowy Jork: Routledge, 1988.
- [8] COHEN B. Explaining psychological statistics. Wyd. 3. Nowy Jork: John Wiley & Sons, 2008.
- [9] BORENSTEIN M. Effect sizes for continuous data. [W:] H. Cooper, L.V. Hedges, J.C. Valentine (red.), The handbook of research synthesis and meta analysis, Nowy Jork, Russell Sage Foundation, 2009, s. 221-237.
- [10] LENHARD W., LENHARD A. Calculation of Effect Sizes. [online:] https://www.psychometrica.de/effect_size.html. Dettelbach (Niemcy): Psychometrica, 2016, doi: 10.13140/RG.2.1.3478.4245.
- [11] BREIMAN L. i in. Classification and regression Trees. Belmont, CA: Wadsworth, 1984.
- [12] IBM SPSS Decision Trees 21. [online:] http://www.sussex.ac.uk/its/pdfs/SPSS_Decision_Trees_21.pdf [dostęp: 22.10.2022].

Opracowano i wydano na podstawie wyników V etapu programu wieloletniego pn. „Poprawa bezpieczeństwa i warunków pracy”, finansowanego w latach 2020-2022 w zakresie badań naukowych i prac rozwojowych ze środków Narodowego Centrum Badań i Rozwoju (projekt IV.PB.01.pt. „Opracowanie narzędzia komputerowego wspomagającego ocenę prawdopodobieństwa powstania wypadku przy pracy i przewidywania jego ciężkości”). Koordynator programu: Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy.