

Załącznik nr 37. Wykaz szczegółowych procedur przygotowujących zbiory KRUS i ZUS

Praca badawcza pt. „Opracowanie metodologii i oszacowanie liczby pracujących w gospodarce narodowej według miejsca zamieszkania i głównego miejsca pracy na poziomie NTS 4, stopy bezrobocia rejestrowanego na poziomie NTS 5 oraz miar wynagrodzeń brutto na poziomie NTS 4”

Wykaz procedur przygotowujących zbiory KRUS

Opis prac przeprowadzonych na zbiorach KRUS w celu zidentyfikowania pracujących i ustalenia pracy głównej:

1. Wyeliminowanie ze zbioru KRUS osób ubezpieczonych, których płatnik nie wystąpił w zbiorze.
Zgodnie z informacjami uzyskanymi od gestora, jeżeli dla danego ubezpieczonego w zbiorze KRUS nie wystąpił płatnik, to takich ubezpieczonych nie należało brać pod uwagę przy dalszym przetwarzaniu zbioru.
2. Aktualizacja kodów TERYT gminy adresu zamieszkania i adresu prowadzenia działalności:
 - a) jeżeli kod TERYT gminy przyjmował wartość niezgodną z rejestrem TERYT ze stanem na 31.12.2016 roku, to wartość tę zastąpiono odpowiednią wartością zgodną z tym rejestrem. Jeśli zmiana dotyczyła rozbicia gminy na część miejską i wiejską, to dla adresów, dla których nazwa gminy była taka sama jak nazwa miejscowości przyjęto symbol części miejskiej, a dla pozostałych symbol części wiejskiej;
 - b) jeżeli kod TERYT gminy przyjmował wartość odpowiadającą delegaturze, to został on zastąpiony kodem TERYT gminy nadrzędnej dla tej delegatury.

3. Weryfikacja poprawności adresów zamieszkania i prowadzenia działalności.

Adres został uznany za poprawny, jeżeli spełnione były poniższe warunki:

- a) kod TERYT gminy był wypełniony i jego wartość była zgodna z rejestrem TERYT ze stanem na dzień 31.12.2016 roku,
- b) flaga spójności adresu była wypełniona i świadczyła o poprawności adresu na poziomie gminy lub niżej (flagi spójności dla adresów są nadawane w US Warszawa podczas przygotowywania zbiorów otrzymanych od gestorów).

Nie wszystkie adresy zostały uznane za poprawne. Konieczne było wdrożenie dalszych działań, które zostały przeprowadzone w oparciu o pozostałe zmienne występujące w zbiorach KRUS oraz inne źródła danych, co opisano w dalszych krokach.

4. Uzupelnienie braków numerów PESEL (zmienna *PESEL_KRUS*) wewnątrz zbioru KRUS.
W pierwszej kolejności utworzono nową zmienną identyfikującą daną osobę. W tym celu wykorzystane zostały zmienne zawierające następujące informacje: nazwisko, imię, data urodzenia, adres zamieszkania (kod TERYT gminy, nazwa miejscowości, przedrostek ulicy, nazwa ulicy, numer budynku i numer lokalu). Warunkiem utworzenia zmiennej było przyjęcie niepustej wartości przez powyższe zmienne (w przypadku danych adresowych warunek ten dotyczył tylko kodu TERYT gminy, nazwy miejscowości i numeru budynku) oraz poprawność adresu zamieszkania. Jeżeli warunek ten nie był spełniony, to zmienna identyfikująca nie była tworzona.

Rekordy z nieuzupełnionym numerem PESEL mogły dotyczyć osób, dla których w zbiorze KRUS występował więcej niż jeden rekord i w co najmniej jednym rekordzie miały one uzupełniony numer PESEL. Wykorzystując nowo utworzoną zmienną identyfikującą porównano jej wartości dla wszystkich rekordów ze zbioru KRUS. W sytuacji, gdy dla rekordu z niewypełnionym numerem PESEL wartość zmiennej identyfikującej odpowiadała wartości zmiennej identyfikującej dla rekordu z wypełnionym numerem PESEL można było dokonać uzupełnienia numeru PESEL.

Po wykonaniu powyższej procedury wykorzystującej metodę imputacji prostej (dedukcyjnej) dla części osób numery PESEL pozostały nieuzupełnione.

5. Uzupelnienie braków numerów PESEL na podstawie informacji ze zbioru podatników CRP KEP, którego gestorem jest Ministerstwo Finansów.

Zbiór CRP KEP zawiera dane identyfikujące podatników, ponadto każdy podatnik występuje w nim tylko jeden raz. W celu połączenia zbioru KRUS ze zbiorem CRP KEP, w zbiorze CRP KEP również należało zweryfikować poprawność adresu zamieszkania oraz utworzyć zmienną identyfikującą osobę - analogicznie jak w zbiorze KRUS. Na potrzeby niniejszej pracy przyjęto, że jeżeli nowa zmienna identyfikowała w zbiorze CRP KEP dokładnie jedną osobę, to na jej podstawie można było uzupełnić numer PESEL w zbiorze KRUS.

Po przeprowadzeniu powyższej procedury część numerów PESEL pozostała nadal nieuzupełniona.

6. Weryfikacja informacji o płci i dacie urodzenia.

Zmienne zawierające informacje o płci i dacie urodzenia (*plec_KRUS* i *data_ur_KRUS*) przyjmowały niepuste wartości w całym zbiorze i nie wymagały uzupełnienia. Weryfikacja wartości zmiennych była możliwa, ponieważ płeć i data urodzenia są zakodowane w numerze PESEL. Została ona dokonana dla rekordów, dla których wartość zmiennej *PESEL_KRUS* była ustalona.

Zmienną odnoszącą się do płci zweryfikowano i zmodyfikowano następująco:

- a) jeżeli zmienna opisująca płeć (zmienna *plec_KRUS*) była wypełniona oraz zmienna zawierająca numer PESEL (zmienna *PESEL_KRUS*) była wypełniona, a informacja o płci zawarta na 10. (przedostatniej) pozycji numeru PESEL była zgodna z informacją o płci zawartą w zmiennej *plec_KRUS*, to informację o płci pobrano ze zmiennej *plec_KRUS*,
- b) w przeciwnym wypadku informację o płci pobrano z numeru PESEL.

Zweryfikowana i zmodyfikowana została również zmienna odnosząca się do daty urodzenia:

- a) jeżeli zmienna opisująca datę urodzenia (zmienna *data_ur_KRUS*) była wypełniona oraz zmienna zawierająca numer PESEL (zmienna *PESEL_KRUS*) była wypełniona, a informacja o dacie urodzenia zawarta w numerze PESEL była zgodna z informacją o dacie urodzenia zawartą w zmiennej *data_ur_KRUS*, to informację o dacie urodzenia pobrano ze zmiennej *data_ur_KRUS*,
- b) w przeciwnym wypadku informację o dacie urodzenia pobrano z numeru PESEL.

7. Utworzenie zmiennej opisującej wiek osoby w dniu referencyjnym badania.

Dla wszystkich rekordów z wypełnioną zmienną opisującą datę urodzenia utworzono zmienną *wiek_KRUS* opisującą wiek osoby w następujący sposób:

- a) jeżeli w okresie od początku roku, w którym przeprowadzono badanie do dnia referencyjnego badania osoba miała urodziny, to jej wiek stanowiła różnica między rokiem referencyjnym badania a rokiem urodzenia,
- b) jeżeli w okresie od początku roku, w którym przeprowadzono badanie do dnia referencyjnego badania osoba nie miała urodzin, to jej wiek stanowiła różnica między rokiem referencyjnym badania a rokiem urodzenia pomniejszona o 1 rok.

8. Ustalenie dla każdego rekordu ze zbioru KRUS czy zawiera on informacje o osobie pracującej według stanu na dzień referencyjny badania.

Ustalenie informacji o fakcie wykonywania pracy według KRUS na dzień referencyjny badania, zgodnie z procedurą opisaną w podrozdziale 2.3.1., było możliwe przy pomocy zmiennych dotyczących:

- statusu osoby ubezpieczonej (*status_KRUS*),
- numeru konta płatnika składek (*nkp_KRUS*),
- wskaźnika podlegania ubezpieczeniu emerytalno-rentowemu (*wsk_podl_uber_KRUS*),
- wskaźnika podlegania ubezpieczeniu wypadkowemu, chorobowemu i macierzyńskiemu (*wsk_podl_ubwcm_KRUS*),
- liczby gospodarstw na koncie płatnika (*liczba_gosp_rol_KRUS*),
- liczby działów specjalnych na koncie płatnika (*liczba_dzial_spec_KRUS*),

- daty rozpoczęcia działalności rolniczej przez płatnika (*data_rozp_dzial_rol_KRUS*),
- daty zgonu płatnika lub osoby ubezpieczonej (*data_zgon_KRUS*),
- wieku płatnika lub osoby ubezpieczonej (*WIEK_2016*),
- wskaźnika pobierania emerytury/renty (*wsk_uber_KRUS*).

Wyodrębnienie podzbioru pracujących ze zbioru KRUS zostało wykonane w poniższych krokach:

- a) w pierwszym etapie wybrano osoby, które podlegały ubezpieczeniu w KRUS na wskazany dzień (zmienna *status_KRUS* przyjęła wartość „A”) lub były płatnikami, którzy nie podlegali ubezpieczeniu na wskazany dzień, ale na ich koncie była zarejestrowana osoba, która podlegała ubezpieczeniu na wskazany dzień (zmienna *wsk_platnik_KRUS* przyjęła wartość „T”, zmienna *status_KRUS* przyjęła wartość „N” lub „Z” i istniała co najmniej jedna osoba ubezpieczona na koncie tego płatnika, dla której zmienna *status_KRUS* przyjmowała wartość „A”);
- b) następnym krokiem było wyłonienie, z wybranej w pierwszym etapie grupy rekordów, wszystkich osób, które były płatnikami oraz osób, które nie były płatnikami, ale podlegały ubezpieczeniu emerytalno-rentowemu na dzień referencyjny badania i jednocześnie podlegały ubezpieczeniu wypadkowemu, chorobowemu i macierzyńskiemu na ten dzień;
- c) w kolejnym kroku dokonano wyboru wszystkich osób, które były płatnikami i posiadały w dniu referencyjnym badania na swoim koncie płatnika co najmniej jedno gospodarstwo lub co najmniej jeden dział specjalny oraz osób niebędących płatnikami, których płatnicy posiadali na swoim koncie płatnika co najmniej jedno gospodarstwo lub co najmniej jeden dział specjalny;
- d) w dalszej kolejności przeanalizowano daty rozpoczęcia działalności rolniczej dla płatników; do kolejnego etapu zakwalifikowano płatników, dla których data rozpoczęcia działalności rolniczej była równa lub wcześniejsza od dnia referencyjnego badania lub była ona niewypełniona oraz osoby ubezpieczone na kontach tych płatników;
- e) następnie wybrano osoby, które nie miały wypełnionej informacji o dacie zgonu lub jeżeli miały ją wypełnioną, to data ta była późniejsza niż dzień referencyjny badania;
- f) zgodnie z przyjętą definicją osób, które nie ukończyły 16 lat nie można było uznać za pracujące; ze zbioru utworzonego w poprzednich krokach wybrano tylko osoby, które w dniu referencyjnym badania miały ukończone 16 lat;
- g) ostatnim krokiem był wybór osób, które w dniu referencyjnym badania nie pobierały renty strukturalnej.

Do zbioru KRUS wprowadzono nową zmienną *pracujący_KRUS*, która dla rekordów wybranych w powyższej procedurze przyjęła wartość „1”, a dla pozostałych rekordów przyjęła wartość „0”. Za pracujących według KRUS uznano osoby, dla których w przynajmniej jednym rekordzie zmienna *pracujący_KRUS* przyjęła wartość „1”.

Przeprowadzenie powyższej procedury zakończyło proces wyboru pracujących według KRUS. Nie było konieczności wdrożenia dalszych działań.

9. Uzupelnienie adresu zamieszkania dla osób uznanych za pracujące wg KRUS.

Adres zamieszkania został uzupełniony w następujących krokach (krok b) był wykonywany, o ile po wykonaniu kroku a) w zbiorze KRUS pozostały jeszcze niepoprawne adresy zamieszkania, itd.):

- a) uzupełnienie adresu zamieszkania danymi z innego rekordu zbioru KRUS dla tej samej osoby (ten sam numer PESEL) - jeżeli dana osoba występowała w zbiorze KRUS w kilku rekordach i w co najmniej jednym z nich miała niewypełniony lub nieprawidłowy adres zamieszkania oraz jednocześnie w przynajmniej jednym z jej rekordów adres zamieszkania był wypełniony i poprawny, to prawidłowy adres posłużył do uzupełnienia zmiennych adresowych w rekordach z niepoprawnym adresem zamieszkania dla tej osoby,
- b) dopisanie adresu zamieszkania ze zbioru CRP KEP - zmiennie dotyczące adresu zamieszkania uzupełniono łącząc zbiór KRUS ze zbiorem CRP KEP przy pomocy zmiennej zawierającej informację o numerze PESEL, a dla osób prowadzących działalność gospodarczą także przy pomocy zmiennej zawierającej informację o numerze NIP,

- c) uzupełnienie adresu zamieszkania z innego rekordu zbioru KRUS dla tego samego numeru konta płatnika (wykonywane tylko dla pracujących płatników) - jeżeli dla danej wartości zmiennej *nkp_KRUS* występował drugi płatnik i adres zamieszkania tego płatnika był poprawny, to korzystano z cech adresowych drugiego płatnika, jeśli drugi płatnik nie występował, ale występowały osoby ubezpieczone na koncie płatnika (z poprawnymi cechami adresowymi), to korzystano z cech adresowych ubezpieczonego najbliższej spokrewnionej z płatnikiem,
- d) uzupełnienie adresu zamieszkania wartościami zmiennych adresu działalności z tego samego rekordu zbioru KRUS, o ile adres działalności był poprawny (wykonywane tylko dla pracujących płatników) - na potrzeby badania przyjęto, że o ile cechy adresowe nie stanowiły inaczej, adres zamieszkania i adres działalności pracującego płatnika znajdował się w tej samej gminie (wśród rekordów pracujących płatników z poprawnym adresem zamieszkania i adresem działalności, rekordy, dla których gmina adresu zamieszkania była taka sama jak gmina adresu działalności stanowiły ponad 89% rekordów).

Po wykonaniu powyższych kroków nie wszyscy pracujący ze zbioru KRUS mieli poprawny adres zamieszkania. Konieczne było dalsze uzupełnianie lub imputacja adresu zamieszkania. Czynności te zostały przeprowadzone osobno dla zbioru pracujących płatników i zbioru pracujących osób ubezpieczonych.

10. Imputacja adresu zamieszkania dla pracujących płatników.

Do uzupełnienia rekordów pracujących płatników z nieustalonym adresem zamieszkania zdecydowano wykorzystać jedną z metod imputacji braków danych. Mając na uwadze cel pracy badawczej imputacją objęto wyłącznie zmienną zawierającą informację o kodzie TERYT gminy zamieszkania. Informacje na temat powiatów, województw, a także charakterystyki miasto-wieś są zakodowane w kodzie TERYT gminy, natomiast informacje o niższych agregacjach (miejscowość, ulica, itp.) nie były potrzebne do realizacji celu pracy badawczej.

Imputacja brakujących kodów TERYT gminy zamieszkania została przeprowadzona na podstawie zmiennych zawierających informację o numerach oddziału terenowego KRUS i placówki terenowej KRUS. Każdy z oddziałów terenowych KRUS obejmuje swoim zasięgiem mieszkańców z terenu głównie jednego województwa. Każdy z oddziałów posiada inne województwo dominujące, a liczba oddziałów jest taka sama jak liczba województw. Podobnie, każda z placówek terenowych obsługuje mieszkańców z terenu od kilku do kilkunastu gmin. Dla każdej gminy istnieje placówka „dominująca”, w której zarejestrowana jest większość rekordów z danej gminy. Mając informację o placówce terenowej KRUS i oddziale terenowym KRUS, pod który ona podlega można było za pomocą wektora prawdopodobieństwa wystąpienia gminy w danej placówce terenowej imputować brakujące kody TERYT gminy zamieszkania.

Ze względu na niewielki udział braków gminy adresu zamieszkania wśród pracujących płatników (0,00035%) podjęto decyzję o zastosowaniu uproszczonej metody imputacji braków. W zbiorze wszystkich pracujących w KRUS posiadających poprawny adres zamieszkania, dla każdej placówki terenowej wyznaczono gminę, dla której w danej placówce występuje najwięcej rekordów i zgodnie z tym wykazem uzupełniono brakujące kody TERYT gminy zamieszkania dla pracujących płatników ze zbioru KRUS. Jeśli dla danej placówki KRUS wystąpiła więcej niż jedna gmina z najwyższą liczbą rekordów występujących w danej placówce, to do uzupełnienia braków wykorzystano gminę, która w klasyfikacji TERYT występuje najwcześniej (przy uszeregowaniu województw od 02 do 16).

Po wykonaniu powyższych czynności wszyscy pracujący płatnicy ze zbioru KRUS mieli poprawny adres zamieszkania. Natomiast dla pracujących osób, które nie były płatnikami pozostały rekordy z niepoprawnym adresem zamieszkania. Dla tych osób adres zamieszkania został uzupełniony w późniejszym etapie wartościami zmiennych adresu zamieszkania ich płatnika.

11. Uzupełnienie adresu działalności dla płatników uznanych za pracujących wg KRUS.

Po wykonaniu wyżej opisanych kroków wszystkie rekordy pracujących płatników posiadały poprawny adres zamieszkania. Aby umożliwić realizację celu badania również adresy działalności pracujących płatników powinny być poprawne. W zbiorze pracujących płatników KRUS dla około 89% rekordów adres zamieszkania i adres działalności znajdowały się w tej samej gminie. W związku z tym dla niepoprawnych adresów działalności za gminę prowadzenia działalności zdecydowano przyjąć gminę adresu zamieszkania.

12. Wprowadzenie identyfikatora osoby pracującej.

Aby umożliwić ustalenie rzeczywistej liczby pracujących w osobach, każdy pracujący powinien mieć unikalną zmienną identyfikującą. Przyjęto, że osoby z niewypełnioną zmienną informującą o numerze PESEL nie dublowały się z osobami z wypełnioną zmienną informującą o numerze PESEL (były to różne osoby), ponieważ podczas wcześniej wykonanych etapów prac uzupełniono numer PESEL na podstawie innych zmiennych występujących w KRUS (tam gdzie było to możliwe).

Do identyfikacji osób nieposiadających wypełnionego numeru PESEL można było użyć m.in. danych osobowych (nazwisko, imię, płeć, data urodzenia) i cech adresowych adresu zamieszkania. Można było zastosować takie rozwiązanie, ponieważ dla wszystkich rekordów z niewypełnioną informacją o numerze PESEL zmienne dotyczące danych osobowych były wypełnione, a adres zamieszkania był poprawny. Wprowadzono nową zmienną *id_osoby*, weryfikującą liczbę wystąpień (rekordów) tej samej osoby w zbiorze pracujących według KRUS. Do jej utworzenia posłużyły zmienne opisujące:

- a) w przypadku osób z wypełnioną zmienną zawierającą numer PESEL – numer PESEL,
- b) w przypadku osób z niewypełnioną zmienną zawierającą numer PESEL - nazwisko, imię, data urodzenia, adres zamieszkania (kod TERYT gminy, nazwa miejscowości, przedrostek ulicy, nazwa ulicy, numer budynku i numer lokalu).

Zmienna *id_osoby* została uzupełniona dla wszystkich rekordów osób pracujących.

13. Wybór pracy głównej.

Praca główna została ustalona tylko dla tych osób zarejestrowanych w zbiorze KRUS, które zostały uznane za pracujące. Jako identyfikator jednoznacznie wskazujący osobę wykorzystana została zmienna *id_osoby*.

Jeżeli dla danej osoby, identyfikowanej na podstawie zmiennej *id_osoby*, w zbiorze pracujących w KRUS wystąpił tylko jeden rekord, to przyjęto, że rekord ten odnosi się do jej pracy głównej. Jeżeli natomiast osoba wystąpiła w zbiorze pracujących w KRUS w więcej niż jednym rekordzie, to mogła dla niej wystąpić jedna z poniższych sytuacji:

- a) osoba wystąpiła w zbiorze KRUS tylko jako płatnik i:
 - pracowała na rzecz co najmniej jednego działu specjalnego,
 - pracowała na rzecz co najmniej jednego gospodarstwa i nie pracowała na rzecz działu specjalnego,
- b) osoba wystąpiła w zbiorze KRUS zarówno jako płatnik jak i ubezpieczony,
- c) osoba wystąpiła w zbiorze KRUS tylko jako ubezpieczony i:
 - wystąpiła tylko jeden raz lub wystąpiła wiele razy, ale z tylko jednym numerem konta płatnika,
 - wystąpiła w zbiorze ubezpieczonych wiele razy z różnymi numerami konta płatnika.

Jeżeli dana osoba wystąpiła w zbiorze KRUS tylko jako płatnik i pracowała na rzecz co najmniej jednego działu specjalnego, to zgodnie z założeniem przyjętym w podrozdziale 2.3.2. uznano, że pracę główną stanowiła praca na rzecz jednego z działów specjalnych. Jeżeli płatnik pracował na rzecz tylko jednego działu specjalnego, to tę pracę uznano za jego pracę główną. W przypadku, gdy pracował na rzecz więcej niż jednego działu specjalnego, wybór pracy głównej został dla niego dokonany w następnym kroku, na podstawie informacji o wielkości gospodarstwa.

Jeżeli dana osoba wystąpiła w zbiorze KRUS tylko jako płatnik i pracowała na rzecz co najmniej jednego gospodarstwa, to za pracę główną dla tej osoby uznano gospodarstwo o największej powierzchni liczonej w hektarach przeliczeniowych.

W przypadkach, gdy nie było można ustalić powierzchni gospodarstwa lub dwa lub więcej gospodarstwa miały taką samą powierzchnię, jako pracę główną wybrano gospodarstwo o najniższym numerze konta płatnika.

Jeżeli istniało więcej niż jedno gospodarstwo przypisane do tego samego konta płatnika, to za pracę główną uznano pracę na rzecz gospodarstwa o najniższym numerze rekordu w zbiorze. W tym celu wykorzystana została wprowadzona przez US Warszawa zmienna *id_wiersza_KRUS*.

Jeżeli dana osoba wystąpiła w zbiorze KRUS zarówno jako płatnik jak i ubezpieczony, to pracę główną dla tej osoby wybrano spośród rekordów, świadczących o byciu płatnikiem, postępując analogicznie jak w przypadku, gdy osoba wystąpiła w zbiorze KRUS wyłącznie jako płatnik.

Jeżeli dana osoba wystąpiła w zbiorze KRUS tylko jako ubezpieczony i wystąpiła w wielu rekordach, ale z tylko jednym numerem konta płatnika, to za jej pracę główną można było przyjąć dowolny rekord, ponieważ cechy ubezpieczonego dostępne w zbiorze KRUS nie miały wpływu na wybór pracy głównej, a cechy płatnika w przypadku takiej osoby były zawsze takie same. Aby w tej sytuacji umożliwić wybór dokładnie jednego rekordu zdecydowano za pracę główną tej osoby uznać rekord o najniższym numerze w zbiorze KRUS.

Jeżeli osoba wystąpiła w zbiorze ubezpieczonych wiele razy z różnymi numerami konta płatnika, to do dalszego etapu wyboru pracy głównej wybrano dla niej najniższy numer konta płatnika. Jeżeli dla najniższego numeru konta płatnika wystąpił więcej niż jeden rekord, to za pracę główną osoby ubezpieczonej uznano rekord o najniższym numerze w zbiorze.

Do zbioru pracujących według KRUS wprowadzono nową zmienną *praca_glowna_KRUS*, która dla rekordów wybranych w powyższej procedurze przyjęła wartość „1”, a dla pozostałych rekordów przyjęła wartość „0”. Dla każdej obserwacji ze zbioru pracujących wg KRUS wystąpił dokładnie jeden rekord, dla którego zmienna *praca_glowna_KRUS* przyjęła wartość „1”. Pracę wykazaną w tym rekordzie uznano za pracę główną według KRUS.

Przeprowadzenie powyższej procedury zakończyło proces wyboru pracy głównej dla każdej osoby pracującej według KRUS. Nie było konieczności wdrożenia dalszych działań.

14. Ustalenie adresu zamieszkania i adresu siedziby pracy głównej dla wszystkich pracujących w pracy głównej.

Aby umożliwić agregację danych po siedzibie pracy głównej dla wszystkich pracujących w pracy głównej należało ustalić siedzibę pracy głównej. Jako adres siedziby pracy przyjęto adres prowadzenia działalności, ale w zbiorze KRUS zmienne dotyczące adresu prowadzenia działalności były wypełnione tylko dla płatników. Dla osób nieposiadających statusu płatnika informacje te nie były wypełnione. W poprzednich krokach ustalono i zweryfikowano adres prowadzenia działalności dla płatników. Należało jeszcze ustalić go dla ubezpieczonych. Zgodnie z ustawą o ubezpieczeniu społecznym rolników ubezpieczeni nieposiadający statusu płatnika są członkami gospodarstwa domowego swojego płatnika i pracują w jego gospodarstwie rolnym. Przyjęto, że dla osób nieposiadających statusu płatnika adresem siedziby pracy głównej będzie adres prowadzenia działalności ich płatnika. W tym celu do zbioru ubezpieczonych posiadających pracę główną dołączono informacje dotyczące płatnika ubezpieczonego.

Cechy płatnika dołączono do ubezpieczonych zgodnie z numerem konta płatnika.

Jeżeli dla danego ubezpieczonego istniał tylko jeden płatnik i dla tego płatnika występował tylko jeden rekord, to zmienne płatnika z tego rekordu dołączono do ubezpieczonego.

Jeżeli dla danego ubezpieczonego istniał więcej niż jeden płatnik lub istniał tylko jeden płatnik i dla tego płatnika wystąpił więcej niż jeden rekord oraz istniał tylko jeden rekord z liczbą ubezpieczonych na koncie płatnika większą lub równą „1”, to zmienne płatnika z tego rekordu dołączono do ubezpieczonego.

Jeżeli dla danego ubezpieczonego istniał więcej niż jeden płatnik lub istniał tylko jeden płatnik i dla tego płatnika wystąpił więcej niż jeden rekord z liczbą ubezpieczonych na koncie płatnika większą lub równą „1”, to wyboru płatnika, którego zmienne dołączono do ubezpieczonego, dokonano spośród tych rekordów.

Jeżeli dla ubezpieczonego istniał jeden lub więcej płatników i wśród rekordów tych płatników znajdował się dokładnie jeden rekord świadczący o pracy głównej, to do ubezpieczonego dołączono zmienne płatnika z tego rekordu.

Jeżeli dla ubezpieczonego istniał jeden lub więcej płatników i wśród rekordów tych płatników wystąpił więcej niż jeden rekord spełniający kryteria pracy głównej, to wyboru płatnika, którego zmienne dołączono do ubezpieczonego, dokonano spośród tych rekordów.

Jeżeli dla ubezpieczonego istniał jeden lub więcej płatników i wśród rekordów tych płatników znajdował się dokładnie jeden rekord świadczący o wykonywaniu pracy przez płatnika, to do ubezpieczonego dołączono zmienne płatnika z tego rekordu.

Jeżeli dla ubezpieczonego istniał jeden lub więcej płatników i wśród rekordów tych płatników wystąpił więcej niż jeden rekord spełniający kryteria wykonywania pracy przez płatnika, to wyboru płatnika, którego zmienne dołączono do ubezpieczonego dokonano spośród tych rekordów.

Jeżeli po wykonaniu powyższych kroków nie udało się dla ubezpieczonego ustalić jednoznacznie rekordu płatnika, to do ubezpieczonego dołączono zmienne płatnika z rekordu o najniższym numerze w zbiorze KRUS.

Po wykonaniu powyższych czynności wszyscy pracujący mieli uzupełniony i poprawny adres siedziby pracy głównej.

15. Uzupełnienie adresu zamieszkania dla pracujących nie posiadających statusu płatnika.

Dla części pracujących ubezpieczonych nie posiadających statusu płatnika nie udało się uzupełnić adresu zamieszkania w poprzednich etapach prac. Można było tego dokonać po dopisaniu informacji dotyczących płatnika do rekordu ubezpieczonego. Za adres zamieszkania ubezpieczonego przyjęto adres zamieszkania jego płatnika.

16. Zapewnienie porównywalności obserwacji w zbiorach KRUS i ZUS.

Aby móc zrealizować cel pracy badawczej, tj. wyodrębnienie pracujących w pracy głównej na bazie zbiorów ZUS i KRUS, dane ze zbioru KRUS należało porównać z danymi ze zbioru ZUS i w przypadku, gdy dana osoba znajdowała się zarówno w zbiorze KRUS jak i w zbiorze ZUS zdecydować, która z prac będzie jej pracą główną. Do dokonania tego porównania niezbędny był identyfikator osoby pozwalający jednoznacznie wskazać rekordy dotyczące tej samej osoby w obu zbiorach. W zbiorze KRUS wprowadzono wcześniej zmienną *id_osoby*, utworzoną na bazie numeru PESEL o ile dla danej osoby udało się ustalić numer PESEL, a w przeciwnym wypadku na bazie zmiennych dotyczących m.in. danych osobowych (nazwisko, imię, płeć, data urodzenia) i cech adresowych adresu zamieszkania. W zbiorach ZUS również została wprowadzona zmienna identyfikująca osobę – *ident_osoby*. Utworzono ją, podobnie jak w KRUS, na bazie numeru PESEL oraz danych osobowych i cech adresu zamieszkania, ale nie korzystano z nazwiska i imienia osoby, ponieważ dane te nie zostały udostępnione przez gestora. Wprowadzono nową zmienną identyfikującą osobę, utworzoną na bazie zmiennych wspólnych dla obu zbiorów. Ponieważ wszystkie zmienne wykorzystane do utworzenia identyfikatora osoby w ZUS występowały również w KRUS zdecydowano, że do zbioru KRUS zostanie dodana nowa zmienna identyfikująca osobę utworzona w ten sam sposób jak zmienna *ident_osoby* w zbiorze ZUS. W zbiorze KRUS utworzono zmienną *id* w następujący sposób:

- a) jeżeli dla danej osoby zmienna zawierająca numer PESEL była wypełniona, to jako zmienną *id* przyjęto wartość zmiennej zawierającej numer PESEL,
- b) jeżeli dla danej osoby zmienna zawierająca numer PESEL nie była wypełniona, to zmienna *id* została utworzona w oparciu o zmienne zawierające rok urodzenia, płeć, adres zamieszkania (kod TERYT gminy, nazwa miejscowości, nazwa ulicy, numer budynku, numer lokalu), o ile kombinacja tych zmiennych była unikalna w zbiorze KRUS,
- c) w pozostałych przypadkach, aby uniknąć przypisania dwóm różnym osobom tej samej wartości zmiennej *id*, wartość zmiennej *id* uzależniono od numeru rekordu w zbiorze KRUS, czyli wartości zmiennej *id_wiersza_KRUS*.

Zmienna *id* przyjęła unikalne wartości dla każdej obserwacji w zbiorze KRUS i zapewniła porównywalność osób pracujących w zbiorach KRUS i ZUS.

17. Przygotowanie zmiennych niezbędnych do naliczenia wskaźników i danych wynikowych określonych celem pracy badawczej.

Do naliczenia wskaźników wynikowych przygotowano następujące zmienne:

- a) identyfikacyjne:
 - *id* – zmienna identyfikująca osobę w zbiorze KRUS,
 - *id_wiersza_KRUS* – zmienna porządkowa identyfikująca rekord w zbiorze KRUS,

b) osobowe:

- *plec_KRUS* – zmienna zawierająca informację o płci osoby,
- *wiek_KRUS* – zmienna zawierająca informację o wieku osoby,
- *wsk_platnik_KRUS* – zmienna zawierająca informację o fakcie bycia płatnikiem KRUS,
- *status_KRUS* – zmienna zawierająca informację o statusie osoby w KRUS,

c) adresowe:

- *adr_zam_kod_gmn_KRUS* – zmienna zawierająca informację o kodzie gminy zamieszkania osoby,
- *adr_mpd_kod_gmn_KRUS* – zmienna zawierająca informację o kodzie gminy miejsca siedziby pracy głównej osoby.

Wśród wartości wszystkich wymienionych wyżej zmiennych nie wystąpiły braki danych.

Opisane postępowanie zakończyło proces przygotowania zbiorów KRUS. Nie było konieczności wdrożenia dalszych działań.

Wykaz procedur przygotowujących zbiory ZUS

Opis prac przeprowadzonych na zbiorach ZUS w celu wybrania zbioru pracujących, bezrobotnych, zatrudnionych oraz ustalenia pracy głównej i wynagrodzeń:

- 1) połączenie dwóch zbiorów: zbioru dotyczącego ubezpieczonych przez płatników fizycznych i zbioru dotyczącego ubezpieczonych przez płatników prawnych lub będących jednostkami organizacyjnymi nieposiadającymi osobowości prawnej.

Połączenie to miało na celu umożliwienie przeprowadzenia wszystkich kolejnych kroków na jednym zbiorze. Znaczna część zmiennych występowała zarówno w zbiorze dotyczących płatników fizycznych jak i płatników prawnych (za okresy wrzesień 2016-czerwiec 2017: 124 zmienne w zbiorze płatników fizycznych i 121 zmiennych w zbiorze płatników prawnych, a 118 zmiennych wystąpiło w obu zbiorach; za okresy marzec 2016 – czerwiec 2016: 102 zmienne w zbiorze płatników fizycznych i 97 zmiennych w zbiorze płatników prawnych, a 96 zmiennych wystąpiło w obu zbiorach).

W celu dokonania połączenia należało:

- a) porównać struktury obu zbiorów oraz zidentyfikować różnice w ich strukturze,
- b) ujednolicić struktury obu zbiorów poprzez utworzenie struktury wspólnej dla obu zbiorów,
- c) połączyć wszystkich ubezpieczonych w ZUS w jeden zbiór,
- d) wprowadzić zmienną *LP* zawierającą informację o numerze rekordu w zbiorze.

Przeprowadzenie powyższej procedury zakończyło proces łączenia zbiorów ZUS. Nie było konieczności wdrożenia dalszych działań.

- 2) uzupełnienie braków numeru PESEL ubezpieczonego numerem PESEL płatnika w przypadku, gdy jednocześnie spełnione były następujące warunki:

- a) osoba ubezpieczona w ZUS prowadziła pozarolniczą działalność gospodarczą i ubezpieczała samą siebie,
- b) liczba osób objętych ubezpieczeniem, za które płatnik rozliczył należne składki, była równa 1,
- c) numer PESEL płatnika nie był brakiem danych,
- d) płeć płatnika była równa płci ubezpieczonego, a płeć płatnika zdefiniowano następująco:
 - jeżeli PESEL płatnika nie był brakiem danych, to zmienną tę uzupełniono na podstawie informacji o płci z 10. (przedostatniej) cyfry numeru PESEL płatnika,
 - jeżeli PESEL płatnika był brakiem danych, to zmienną tę pozostawiono niewypełnioną,
- e) rok urodzenia płatnika był tożsamy z rokiem urodzenia ubezpieczonego,
- f) kod TERYT gminy zamieszkania płatnika był równy kodowi TERYT gminy zamieszkania ubezpieczonego.

Dla rekordów spełniających powyższe warunki numer PESEL ubezpieczonego uzupełniono numerem PESEL płatnika.

Ze względu na nadal występujące braki w numerze PESEL ubezpieczonego istniała konieczność wdrożenia dalszych działań, które zostały przeprowadzone na podstawie pozostałych zmiennych występujących w zbiorach ZUS oraz innych źródeł danych, co opisano w dalszych krokach.

- 3) aktualizacja, zmiana oraz sprawdzenie poprawności kodów gminy na podstawie odpowiedniego rejestru TERYT.

W związku ze zmieniającym typem niektórych gmin konieczne było zaktualizowanie kodów TERYT gminy o zmiany, które zaszły od roku 2013. Zastąpione zostały także kody gmin w przypadkach delegatur oraz dzielnic Warszawy oraz dopisane informacje o miastach i częściach wiejskich w gminach miejsko-wiejskich, w których nie było informacji o podziale na miasto i wieś.

Kod gminy został uznany za poprawny, gdy znajdował się w odpowiednim rejestrze TERYT, miał wypełnioną informację o fladze spójności i informacja ta wskazywała na poprawność adresu.

- 4) uzupełnienie adresu zamieszkania ubezpieczonego adresem zameldowania ubezpieczonego (zgodnie z zasadami przyjętymi w ZUS adres zamieszkania jest wypełniony, jedynie gdy jest on różny od adresu zameldowania).

Zmienne dotyczące adresu zamieszkania ubezpieczonego uzupełniono wartościami zmiennych dotyczących adresu zameldowania ubezpieczonego, jeśli wszystkie zmienne dotyczące adresu zamieszkania były brakami danych oraz adres zameldowania został uznany za poprawny.

Powyższe zmienne przyjęły wartości analogicznych zmiennych występujących w adresie zameldowania.

Ze względu na nadal występujące braki w zmiennych dotyczących adresu zamieszkania ubezpieczonego konieczne było wdrożenie dalszych działań, które zostały przeprowadzone w oparciu o pozostałe zmienne występujące w zbiorach ZUS oraz inne źródła danych, co opisano w dalszych krokach.

- 5) utworzenie zmiennej identyfikującej ubezpieczonego (*ident_osoby*).

- a) jeżeli adres zamieszkania ubezpieczonego został uznany za poprawny oraz wartości zmiennych opisujących rok urodzenia, płeć, nazwę miejscowości, kod TERYT gminy i numer budynku zamieszkania ubezpieczonego nie były brakiem danych, to zmienna *ident_osoby* została uzupełniona ciągiem znaków, na który składały się: rok urodzenia, płeć, adres zamieszkania (kod TERYT gminy, nazwa miejscowości, przedrostek ulicy, nazwa ulicy, numer budynku i numer lokalu zamieszkania ubezpieczonego),

- b) jeżeli adres zamieszkania ubezpieczonego został uznany za niepoprawny lub dla przynajmniej jednej ze zmiennych opisujących rok urodzenia, płeć, nazwę miejscowości, kod TERYT gminy i numer budynku jej wartość była brakiem danych, to zmiennej *ident_osoby* przypisano ciąg znaków: „LP”, numer rekordu w zbiorze (zmienna LP), „ZUS”.

Utworzenie zmiennej pozwoliło na przeprowadzenie kolejnych procedur na zbiorze ZUS.

- 6) uzupełnienie braków numeru PESEL ubezpieczonego za pomocą zmiennej *ident_osoby*.

Brak wartości w zmiennej dotyczącej numeru PESEL osoby ubezpieczonej w ZUS uzupełniono przy pomocy zmiennej *ident_osoby*, w przypadku, gdy jednocześnie w bazie występował rekord o tej samej wartości zmiennej *ident_osoby* z wypełnionym numerem PESEL. Brak uzupełniono numerem PESEL z tego rekordu.

Ze względu na nadal występujące braki w numerze PESEL ubezpieczonego konieczne było wdrożenie dalszych działań, które zostały przeprowadzone przy pomocy zbioru KEP, co opisano w kolejnym kroku.

- 7) połączenie zbioru ZUS ze zbiorem KEP oraz uzupełnienie numeru PESEL w oparciu o zmienne pochodzące ze zbioru KEP.

W zbiorze KEP utworzono zmienną *ident_osoby_KEP_dla_ZUS* na podstawie takich zmiennych jakie wykorzystano do utworzenia zmiennej *ident_osoby* w zbiorze ZUS, tzn.:

- a) rok urodzenia (ustalany na podstawie czterech pierwszych cyfr zmiennej *data_ur_KEP*),
- b) płeć (ustalaną na podstawie przedostatniej cyfry numeru PESEL zawartego w zmiennej *PESEL_KEP*),

c) zmienne adresowe: *adr_zam_kod_gmn_KEP*, *adr_zam_naz_msc_KEP*, *adr_zam_pr_ulica_KEP*, *adr_zam_naz_ulica_KEP*, *adr_zam_nr_bud_KEP*, *adr_zam_nr_lok_KEP*.

Jeżeli zmienna *ident_osoby* ze zbioru ZUS przyjęła taką samą wartość jak zmienna *ident_osoby_KEP_dla_ZUS* ze zbioru KEP, to dla danego rekordu w zbiorze ZUS dopisano następujące zmienne ze zbioru KEP:

- a) *ident_osoby_KEP_dla_ZUS*,
- b) *COUNT_of_ident_osoby_KEP_dla_ZUS*,
- c) *KEP*,
- d) *id_wiersza_KEP*,
- e) *nazwisko_KEP*,
- f) *imie_KEP*,
- g) *PESEL_KEP*,
- h) *data_ur_KEP*,
- i) *msc_ur_KEP*,
- j) *kod_kraj_ur_KEP*,
- k) *adr_zam_kod_kraj_KEP*,
- l) *adr_zam_kod_gmn_KEP*,
- m) *adr_zam_naz_msc_KEP*,
- n) *adr_zam_naz_ulica_KEP*,
- o) *adr_zam_pr_ulica_KEP*,
- p) *adr_zam_nr_bud_KEP*,
- q) *adr_zam_nr_lok_KEP*,
- r) *adr_zam_tc_KEP*,
- s) *adr_zam_kod_pocz_KEP*,
- t) *adr_zam_poczta_KEP*.

Uzupełnienia braku numeru PESEL ubezpieczonego dokonano w przypadkach, gdy dla danego rekordu zmienna *ident_osoby* przyjęła taką samą wartość jak zmienna *ident_osoby_KEP_dla_ZUS*, wartość zmiennej *PESEL_KEP* nie była brakiem danych oraz gdy dana wartość zmiennej *ident_osoby_KEP_dla_ZUS* wystąpiła tylko raz w zbiorze KEP.

Po przeprowadzeniu powyższej procedury część numerów PESEL pozostała nadal nieuzupełniona.

- 8) weryfikacja i uzupełnienie informacji o płci i dacie urodzenia na podstawie numeru PESEL.

Weryfikacji można było dokonać, ponieważ płeć i data urodzenia są zakodowane w numerze PESEL.

Dla wszystkich rekordów z wypełnioną zmienną dotyczącą numeru PESEL zweryfikowano i zmodyfikowano zmienną odnoszącą się do płci w następujący sposób:

- a) jeżeli zmienna opisująca płeć (zmienna *ubezp_plec_ZUS*) była wypełniona oraz numer PESEL był wypełniony, a informacja o płci zawarta na 10. (przedostatniej) pozycji numeru PESEL nie była zgodna z informacją o płci zawartą w zmiennej *ubezp_plec_ZUS*, to informację o płci pobrano z numeru PESEL,
- b) jeżeli zmienna opisująca płeć (zmienna *ubezp_plec_ZUS*) była wypełniona oraz numer PESEL był wypełniony i informacja o płci zawarta na 10. (przedostatniej) pozycji numeru PESEL była zgodna z informacją o płci zawartą w zmiennej *ubezp_plec_ZUS*, to informację o płci pobrano ze zmiennej *ubezp_plec_ZUS*,
- c) jeżeli zmienna opisująca płeć (zmienna *ubezp_plec_ZUS*) nie była wypełniona oraz numer PESEL był wypełniony, to informację o płci pobrano z numeru PESEL,

- d) jeżeli zmienna opisująca płeć (zmienna *ubezp_plec_ZUS*) nie była wypełniona oraz numer PESEL nie był wypełniony, to zmienna *ubezp_plec_ZUS* pozostała niewypełniona.

Po przeprowadzeniu powyższej procedury wykorzystującej metodę imputacji prostej (dedukcyjnej) zmienna określająca płeć dla części ubezpieczonych pozostała niewypełniona. Braki wartości w zmiennej *ubezp_plec_ZUS* dla tych osób zostały uzupełnione przy zastosowaniu metody imputacji wielokrotnej na poziomie danych jednostkowych. Proces imputacji został opisany w podrozdziale 2.3.8.

Następnie dla wszystkich rekordów z wypełnionym numerem PESEL zweryfikowana i zmodyfikowana została zmienna odnosząca się do roku urodzenia *ubezp_rok_ur_ZUS*:

- a) jeżeli zmienna opisująca rok urodzenia (zmienna *ubezp_rok_ur_ZUS*) była wypełniona oraz numer PESEL był wypełniony, a informacja o roku urodzenia zawarta w numerze PESEL nie była zgodna z informacją o roku urodzenia zawartą w zmiennej *ubezp_rok_ur_ZUS*, to informację o roku urodzenia pobrano z numeru PESEL,
- b) jeżeli zmienna opisująca rok urodzenia (zmienna *ubezp_rok_ur_ZUS*) była wypełniona oraz numer PESEL był wypełniony i informacja o roku urodzenia zawarta w numerze PESEL była zgodna z informacją o roku urodzenia zawartą w zmiennej *ubezp_rok_ur_ZUS*, to informację o roku urodzenia pobrano ze zmiennej *ubezp_rok_ur_ZUS*,
- c) jeżeli zmienna opisująca rok urodzenia (zmienna *ubezp_rok_ur_ZUS*) nie była wypełniona oraz numer PESEL był wypełniony, to informację o roku urodzenia pobrano z numeru PESEL,
- d) jeżeli zmienna opisująca rok urodzenia (zmienna *ubezp_rok_ur_ZUS*) nie była wypełniona oraz numer PESEL nie był wypełniony, to zmienna *ubezp_rok_ur_ZUS* pozostała niewypełniona.

Po przeprowadzeniu powyższej procedury wykorzystującej metodę imputacji prostej (dedukcyjnej) zmienna określająca rok urodzenia dla części obserwacji pozostała niewypełniona. Ze względu na pomocniczy charakter zmiennej nie było potrzeby dalszego uzupełniania brakujących wartości.

- 9) utworzenie zmiennej opisującej wiek osoby (zmienna *ubezp_wiek*):

- a) jeżeli osoba miała wypełnioną zmienną dotyczącą numeru PESEL (zawierającego m.in. informację o dacie urodzenia) i w okresie od początku roku, w którym przeprowadzono badanie do dnia referencyjnego badania osoba miała urodziny, to za wiek osoby przyjęto różnicę między rokiem przeprowadzenia badania a rokiem urodzenia,
- b) jeżeli osoba miała wypełnioną zmienną dotyczącą numeru PESEL (zawierającego m.in. informację o dacie urodzenia) i w okresie od początku roku, w którym przeprowadzono badanie do dnia referencyjnego badania osoba nie miała urodzin, to za wiek osoby przyjęto różnicę między rokiem przeprowadzenia badania a rokiem urodzenia pomniejszoną o 1 rok,
- c) jeżeli osoba miała niewypełnioną zmienną dotyczącą numeru PESEL, ale miała wypełnioną zmienną zawierającą informację o roku urodzenia, to za jej wiek przyjęto różnicę między rokiem przeprowadzenia badania a rokiem urodzenia.

Po przeprowadzeniu powyższej procedury wykorzystującej metodę imputacji prostej (dedukcyjnej) zmienna określająca wiek dla części ubezpieczonych pozostała niewypełniona. Braki wartości w zmiennej *ubezp_wiek* dla tych osób zostały uzupełnione przy zastosowaniu metody imputacji wielokrotnej na poziomie danych jednostkowych. Proces imputacji został opisany w podrozdziale 2.3.8.

- 10) połączenie zbioru ZUS z Bazą Jednostek Statystycznych w celu uzupełnienia informacji o płatnikach. Poniższa procedura pozwoliła na uzupełnienie informacji o sekcji PKD 2007 płatnika osoby ubezpieczonej. W tym celu należy wykorzystano następujące zbiory BJS:

- a) zbiór zawierający jednostki lokalne,
- b) zbiór zawierający jednostki prawne,
- c) zbiór zawierający rodzaje jednostek.

Do połączenia wykorzystano zmienne zawierające numer REGON oraz NIP. Przy czym, zbiór zawierający jednostki prawne można było dołączyć przy pomocy obu zmiennych, natomiast do

dołączenia zbioru zawierającego jednostki lokalne mogła być wykorzystana tylko zmienna zawierająca numer REGON. Połączenie było możliwe do przeprowadzenia dla płatników posiadających wypełniony numer REGON lub NIP, o ile był on zgodny z numerem REGON lub NIP jednostki występującej w zbiorze BJS.

Przeprowadzenie powyższej procedury nie uzupełniło informacji o sekcji PKD 2007 płatnika osoby ubezpieczonej dla wszystkich ubezpieczonych (przy łączeniu z BJS w każdym z badanych okresów wystąpiło ok. 0,79% rekordów, które się nie sparowały). Konieczne było wdrożenie dalszych działań, które zostały przeprowadzone w oparciu o pozostałe zmienne występujące w zbiorach ZUS oraz inne źródła danych, co opisano w dalszych krokach.

11) uzupełnienie braków w zmiennej dotyczącej sekcji PKD 2007 płatnika osoby ubezpieczonej.

Część braków w zmiennej dotyczącej sekcji PKD 2007 płatnika osoby ubezpieczonej została uzupełniona w poniższy sposób:

- a) przy wykorzystaniu zmiennej zawierającej informację o sekcji PKD 2007 pochodzącej ze zbioru KEP - dokonano połączenia zbioru ZUS ze zbiorami płatników KEP (zbiory zostały połączone za pomocą NIP lub za pomocą numeru REGON płatnika) i jeżeli wartość zmiennej dotyczącej sekcji PKD 2007 dla danego płatnika w zbiorze ZUS była brakiem danych, to zastąpiono ją informacją o sekcji PKD 2007 pochodzącą ze zbioru KEP,
- b) za pomocą kodu tytułu ubezpieczenia (tylko dla wybranych kodów tytułu ubezpieczenia):
 - jeżeli kod tytułu ubezpieczenia zaczynał się od znaków „0310”, „0320”, to zmienna określająca sekcję PKD 2007 przyjęła wartość „A”,
 - jeżeli kod tytułu ubezpieczenia zaczynał się od znaków „0610”, „1110”, „1111”, „1112”, „1113”, „1114”, „1120”, „1130”, „1131”, „1132”, „1133”, „1134”, „1135”, „1140”, „2010”, „2020”, to zmienna określająca sekcję PKD 2007 przyjęła wartość „O”,
 - jeżeli kod tytułu ubezpieczenia zaczynał się od znaków „1010”, „1011”, „1012”, to zmienna określająca sekcję PKD 2007 przyjęła wartość „S”.

Po przeprowadzeniu powyższej procedury zmienna określająca sekcję PKD 2007 dla części płatników pozostała niewypełniona. W celu uzupełnienia pozostałych braków danych zastosowano metody imputacji wielokrotnej na poziomie danych jednostkowych. Proces imputacji został opisany w podrozdziale 2.3.8.

12) uzupełnienie i korekta zmiennych dotyczących adresu zamieszkania ubezpieczonego.

W przypadku gdy zmienne dot. adresu były niewypełnione lub wypełnione nieprawidłowo zastosowano poniższe postępowanie.

Zmienne dotyczące adresu zamieszkania ubezpieczonego uzupełniono lub skorygowano:

- a) za pomocą zmiennych dotyczących adresu zamieszkania płatnika (w przypadku, gdy płatnik ubezpieczał samego siebie):
 - jeżeli płatnik był osobą fizyczną o tym samym numerze PESEL co ubezpieczony oraz adres zamieszkania płatnika został uznany za poprawny, to zmienne dotyczące adresu zamieszkania ubezpieczonego uzupełniono wartościami zmiennych dotyczących adresu zamieszkania płatnika,
- b) za pomocą zmiennych dotyczących adresu zamieszkania podatnika pochodzących ze zbioru KEP:
 - jeżeli numer PESEL podatnika w zbiorze KEP przyjął tę samą wartość co numer PESEL ubezpieczonego oraz adres zamieszkania podatnika został uznany za poprawny, to zmienne dotyczące adresu zamieszkania ubezpieczonego uzupełniono wartościami zmiennych dotyczących adresu zamieszkania podatnika ze zbioru KEP,
 - jeżeli w zbiorze KEP kod kraju zamieszkania podatnika był różny od kodu Polski (Klasyfikacja ISO 3166), to kod gminy adresu zamieszkania ubezpieczonego uzupełniono wartością „XXXXXX” oznaczającą, że ubezpieczony zamieszkiwał poza granicami Polski i adres ten uznano za poprawny,
- c) za pomocą zmiennych dotyczących adresu zamieszkania pochodzących z innego rekordu w zbiorze ZUS dotyczącego tej samej osoby ubezpieczonej:

- jeżeli w zbiorze ZUS istniał inny rekord przyjmujący tę samą wartość zmiennej zawierającej numer PESEL, co numer PESEL ubezpieczonego i adres zamieszkania ubezpieczonego w tym rekordzie został uznany za poprawny, to dla każdego wypełnionego numeru PESEL wybrano rekord z najpóźniejszą datą początku obowiązywania składek spośród rekordów o poprawnym adresie ubezpieczonego i uzupełniono cechami adresowymi z tego rekordu wszystkie rekordy dla tego numeru PESEL.

Przeprowadzenie powyższej procedury nie zakończyło procesu ze względu na nadal występujące braki w adresie zamieszkania ubezpieczonego. Konieczne było wdrożenie dalszych działań, które zostały przeprowadzone w oparciu o pozostałe zmienne występujące w zbiorach ZUS, co opisano w dalszych krokach.

- 13) ustalenie dla każdego rekordu ze zbioru ZUS informacji o fakcie wykonywania pracy przez ubezpieczonego zgodnie z procedurą opisaną w podrozdziale 2.3.1.

Za pracujących według ZUS uznano osoby ubezpieczone, które spełniały jeden z poniższych warunków:

- a) byli ubezpieczeni z kodem deklaracji ZUS różnym od „90” oraz z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki należały do zbioru („0110”, „0111”, „0200”, „0310”, „0320”, „0511”, „0543”, „0545”, „0581”, „0610”, „0800”, „1010”, „1011”, „1012”, „1110”, „1111”, „1112”, „1113”, „1114”, „1120”, „1130”, „1131”, „1132”, „1133”, „1134”, „1135”, „1140”, „1240”, „2010”, „2020”) i jednocześnie nie otrzymywały świadczenia rehabilitacyjnego w ostatnim miesiącu kwartału
 - b) byli ubezpieczeni z kodem deklaracji ZUS różnym od „90” oraz z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki należały do zbioru („0510”, „0512”, „0513”, „0520”, „0530”, „0540”, „0544”, „0570”, „0572”, „0580”) i jednocześnie nie otrzymywały świadczenia rehabilitacyjnego w ostatnim miesiącu kwartału oraz nie zachodziła sytuacja, że ich płatnik ubezpieczał dokładnie jedną osobę, będąc osobą fizyczną prowadzącą działalność gospodarczą zarejestrowaną w Ewidencji Działalności Gospodarczej lub Centralnej Ewidencji i Informacji o Działalności Gospodarczej, która to działalność została zawieszona po dacie początku obowiązywania składek.
 - c) byli ubezpieczeni z kodem deklaracji ZUS różnym od „90” oraz z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki należały do zbioru („0110”, „0111”, „0200”, „0310”, „0320”, „0511”, „0543”, „0545”, „0581”, „0610”, „0800”, „1010”, „1011”, „1012”, „1110”, „1111”, „1112”, „1113”, „1114”, „1120”, „1130”, „1131”, „1132”, „1133”, „1134”, „1135”, „1140”, „1240”, „2010”, „2020”) i otrzymywały świadczenie rehabilitacyjne, które zakończyło się przed ostatnim dniem kwartału (kod świadczenia/przerwy był równy „321” lub „322”),
 - d) byli ubezpieczeni z kodem deklaracji ZUS różnym od „90” oraz z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki należały do zbioru („0510”, „0512”, „0513”, „0520”, „0530”, „0540”, „0544”, „0570”, „0572”, „0580”) i otrzymywały świadczenie rehabilitacyjne, które zakończyło się przed ostatnim dniem kwartału (kod świadczenia/przerwy był równy „321” lub „322”) oraz nie zachodziła sytuacja, że ich płatnik ubezpieczał dokładnie jedną osobę, będąc osobą fizyczną prowadzącą działalność gospodarczą zarejestrowaną w Ewidencji Działalności Gospodarczej lub Centralnej Ewidencji i Informacji o Działalności Gospodarczej, która to działalność została zawieszona po dacie początku obowiązywania składek,
 - e) byli ubezpieczeni z kodem tytułu ubezpieczenia świadczącym o przebywaniu na urlopie wychowawczym (cztery pierwsze znaki kodu tytułu ubezpieczenia były równe „1211”) i jednocześnie urlop ten rozpoczęły nie wcześniej niż w pierwszym dniu badanego kwartału,
- 14) ustalenie dla każdego rekordu ze zbioru ZUS informacji o bezrobociu ubezpieczonego.
Za bezrobotnych według ZUS uznano osoby ubezpieczone, które spełniały poniższe warunki:
 - a) byli ubezpieczeni z kodem tytułu ubezpieczenia „0910” lub „0911”, ponieważ w ten sposób są zgłaszane do ubezpieczeń w ZUS osoby bezrobotne zarejestrowane przez powiatowe urzędy pracy; przyjęto, że nie należy brać pod uwagę osób z wyżej wymienionymi kodami tytułu ubezpieczenia, których płatnikiem nie były powiatowe urzędy pracy,

- b) na dzień referencyjny badania miały ukończony 18. rok życia i nie osiągnęły wieku emerytalnego (na podstawie utworzonej zmiennej *wiek_emerytalny*, która określała czy dana osoba osiągnęła wiek emerytalny w dniu referencyjnym badania),
 - c) data początku obowiązywania składki w ZUS była wcześniejsza lub równa dacie referencyjnej zbioru ZUS; było to konieczne, ponieważ w zbiorze ZUS znalazły się również informacje wprowadzone po dniu referencyjnym zbioru. Jest to efekt cyklu przetwarzania danych w systemie informatycznym ZUS.
- 15) ustalenie, ile razy i z jakimi kodami tytułu ubezpieczenia osoba spełniająca kryteria osoby bezrobotnej wystąpiła w zbiorze ZUS.

Dla każdej osoby występującej w zbiorze ZUS przyznano status osoby bezrobotnej rekordowi, który wybrano w następujący sposób:

- a) jeśli osoba wystąpiła w zbiorze ZUS tylko raz i spełniała warunek bycia ubezpieczonym, to przyjęto, że posiadała ona status osoby bezrobotnej na dzień referencyjny zbioru.
 - b) jeśli osoba wystąpiła w zbiorze ZUS więcej niż raz z kodami tytułu ubezpieczenia osoby bezrobotnej, to należało ustalić, które wystąpienie najlepiej opisywało osobę w dniu referencyjnym badania; status osoby bezrobotnej przyznano, dla rekordu z późniejszą datą początku obowiązywania składki, a w sytuacji gdy data początku obowiązywania składki była w każdym przypadku taka sama, to na potrzeby niniejszego badania przyjęto, że osobę bezrobotną charakteryzował ten rekord w zbiorze, który miał mniejszą liczbę porządkową (zmienna *LP*).
 - c) jeśli osoba wystąpiła w zbiorze ZUS więcej niż raz z kodami tytułu ubezpieczenia osoby bezrobotnej jak i innymi, to przyjęto, że o fakcie posiadania statusu osoby bezrobotnej na dzień referencyjny badania decydowała data początku obowiązywania składki; status osoby bezrobotnej przyznano, kiedy data początku obowiązywania składki dla rekordu z kodem tytułu ubezpieczenia „0910” lub „0911” była późniejsza od daty początku obowiązywania składki dla rekordów z innymi kodami tytułu ubezpieczenia.
- 16) ustalenie informacji o osobach niespełniających warunków zaliczenia do zbiorowości bezrobotnych zarejestrowanych.
- Za osoby niespełniające warunków zaliczenia do zbiorowości bezrobotnych zarejestrowanych uznano osoby, dla których wystąpiła jedna z poniższych sytuacji:
- a) były ubezpieczone z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki były równe „0910” i dla ich płatnika rodzaj jednostki był inny niż powiatowy urząd pracy,
 - b) były ubezpieczone z kodem deklaracji ZUS równym „90” oraz z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki były równe „1240”,
 - c) były ubezpieczone z kodem tytułu ubezpieczenia, którego cztery pierwsze znaki należały do zbioru („0411”, „0417”, „0421”, „0424”, „0426”, „0428”, „0430”, „0431”, „0920”).

- 17) ustalenie dla każdej osoby uznanej za pracującą lub bezrobotną jej statusu ekonomicznego wg ZUS.

Status ekonomiczny wg ZUS ustalono zgodnie z poniższymi regułami:

- a) jeżeli dla danej osoby rekordem z najpóźniejszą datą początku obowiązywania składek był rekord świadczący o pracy, osobę uznawano za pracującą,
 - b) jeżeli dla danej osoby rekordem z najpóźniejszą datą początku obowiązywania składek był rekord świadczący o bezrobociu, osobę uznawano za bezrobotną,
 - c) jeżeli dla danej osoby rekordem z najpóźniejszą datą początku obowiązywania składek był rekord świadczący o wykluczeniu z bezrobocia oraz dla danej osoby istniał rekord świadczący o pracy, to osobę uznawano za pracującą.
- 18) sprawdzenie czy osoba spełniająca kryteria osoby bezrobotnej w zbiorze ZUS, wystąpiła w zbiorze KRUS i została uznana za pracującą.

Takiego porównania należało dokonać, ponieważ ubezpieczone w KRUS osoby pracujące w indywidualnych gospodarstwach rolnych, zgodnie z obowiązującym stanem prawnym mogły zarejestrować się w powiatowym urzędzie pracy jako osoby bezrobotne. W takich przypadkach należało wyeliminować daną osobę ze zbioru osób bezrobotnych, gdyż według metodologii przyjętej w niniejszej pracy uznano je za osoby pracujące.

Powyższe rozwiązania dotyczące statusu na rynku pracy były możliwe do zastosowania na podstawie informacji dotyczących:

- a) kodu tytułu ubezpieczenia osoby ubezpieczonej,
- b) występowania kodu świadczenia/przerwy,
- c) daty końcowej występowania kodu świadczenia/przerwy,
- d) daty początku obowiązywania składek w kontekście płatnika oraz danego tytułu ubezpieczenia,
- e) rodzaju ubezpieczenia,
- f) rodzaju jednostki w BJS,
- g) podstawowej formy prawnej w BJS,
- h) stanu aktywności ekonomicznej w BJS.

Przeprowadzenie powyższej procedury zakończyło proces uzupełniania informacji o fakcie wykonywania pracy wg ZUS lub posiadania statusu bezrobotnego przez ubezpieczonego. Nie było konieczności wdrożenia dalszych działań.

19) braki w kodzie TERYT gminy adresu siedziby płatnika uzupełniono:

- a) za pomocą zmiennych z BJS,
- b) za pomocą zmiennych z KEP,

20) uzupełnienie braków w kodzie województwa siedziby płatnika

Województwo siedziby płatnika uzupełniono na podstawie oddziału ZUS. Było to możliwe, ponieważ dla 99,8% rekordów z wypełnionym województwem siedziby płatnika występowało ono w ramach jednego oddziału ZUS.

Przeprowadzenie powyższej procedury zakończyło proces uzupełniania braków w kodzie województwa adresu siedziby płatnika, natomiast nie zakończyło ono procesu uzupełniania braków w kodzie gminy adresu siedziby płatnika. W celu uzupełnienia pozostałych braków danych zastosowano metody imputacji wielokrotnej na poziomie danych jednostkowych. Proces imputacji został opisany w podrozdziale 2.3.8.

21) wybór pracy głównej według ZUS dla każdej osoby pracującej w ostatnim dniu okresu referencyjnego.

W celu ustalenia liczby pracujących w pracy głównej należało doprowadzić do sytuacji, w której każda osoba ubezpieczona z tytułu świadczenia pracy była liczona dokładnie jeden raz w swojej pracy głównej zgodnie z założeniem przyjętym w podrozdziale 2.3.2.

Dla wszystkich rekordów osób ubezpieczonych z tylko jednym kodem tytułu ubezpieczenia świadczącym o wykonywaniu pracy rekord ten należy uznać za pracę główną.

Dla wszystkich rekordów osób ubezpieczonych z więcej niż jednym kodem tytułu ubezpieczenia świadczącym o wykonywaniu pracy pracę główną wybrano na podstawie wartości zmiennych opisujących:

- a) wymiar czasu pracy,
- b) wysokość podstawy wymiaru składki emerytalno-rentowej,
- c) hierarchię kodów tytułu ubezpieczenia,
- d) datę początku obowiązywania składek,
- e) numer rekordu w zbiorze.

W zbiorach ZUS wymiar czasu pracy określony w umowie o pracę wpisany jest w postaci ułamka zwykłego (licznik/mianownik) np.:

- a) 1/1 - dla pełnego wymiaru czasu pracy,
- b) 1/2 - dla połowy wymiaru czasu pracy,
- c) 3/4 - dla trzech czwartych wymiaru czasu pracy,
- d) 7/8 - dla siedmiu ósmych wymiaru czasu pracy,
- e) 150/176 - gdy wymiar czasu pracy określony jest w godzinach.

Na podstawie tej informacji utworzono zmienną określającą wymiar czasu pracy (zmienna *wymiar_czasu_pracy_ZUS*), zdefiniowaną następująco:

- a) jeżeli 4 pierwsze znaki kodu tytułu ubezpieczenia były równe „0310”, „0320”, „0610”, „1010”, „1011”, „1012”, „1110”, „1111”, „1112”, „1113”, „1114”, „1120”, „1130”, „1131”, „1132”, „1133”, „1134”, „1135”, „1140”, „2010”, „2020”, to zmienna przyjęła wartość „1.00”,
- b) jeżeli 4 pierwsze znaki kodu tytułu ubezpieczenia były równe „0110”, „0111”, „0200”, „0800”, „1240” i jednocześnie licznik był równy mianownikowi, to zmienna przyjęła wartość „1.00”,
- c) jeżeli 4 pierwsze znaki kodu tytułu ubezpieczenia były równe „0110”, „0111”, „0200”, „0800”, „1240”, licznik był mniejszy od mianownika, licznik i mianownik były różne od 0 oraz iloraz licznika i mianownika był większy lub równy 0.01, to zmienna przyjęła wartość ilorazu licznika i mianownika,
- d) w pozostałych przypadkach zmienna przyjęła wartość nieustaloną.

Dla kodów tytułu ubezpieczenia rozpoczynających się od „05” oraz dla kodu „1211” wymiar czasu pracy nie był ustalany.

Na potrzeby ustalenia pracy głównej przyjęto, że jeżeli osoba jest ubezpieczona w ZUS z co najmniej jednym kodem tytułu ubezpieczenia świadczącym o prowadzeniu pozarolniczej działalności gospodarczej (pierwsze dwa znaki kodu tytułu ubezpieczenia są równe „05”), to:

- a) jeżeli wśród wszystkich kodów tytułu ubezpieczenia tej osoby nie wystąpił żaden rekord, dla którego wartość zmiennej *wymiar_czasu_pracy_ZUS* była równa „1.00”, to za pracę główną zostanie przyjęta jedna z prac z kodem „05”, a pozostałe prace dla tej osoby uznano za prace dodatkowe,
- b) jeżeli wśród wszystkich kodów tytułu ubezpieczenia tej osoby wystąpił co najmniej jeden rekord, dla którego wartość zmiennej *wymiar_czasu_pracy_ZUS* była równa „1.00”, to za pracę główną przyjęta będzie jedna z prac, dla których wartość zmiennej *wymiar_czasu_pracy_ZUS* była równa „1.00”, a pozostałe prace dla tej osoby uznano za prace dodatkowe.

Po ustaleniu wymiaru czasu pracy wybrano wszystkie rekordy dla danej osoby, dla których wartości zmiennej *wymiar_czasu_pracy_ZUS* osiągały maksymalną wartość. Rekordy z wartością zmiennej inną niż maksymalna należało pominąć. Jeśli występował tylko jeden rekord z maksymalną wartością zmiennej, to należało uznać go za pracę główną, a jeżeli rekordów z maksymalną wartością zmiennej było więcej, to należało przejść do kolejnego etapu procedury wyboru pracy głównej. Jeśli dana osoba występowała w rejestrze ZUS wyłącznie z nieustaloną wartością wymiaru czasu pracy, również należało przejść do kolejnego etapu procedury wyboru pracy głównej.

W kolejnym etapie należało porównać wartości zmiennych opisujących podstawy składek na ubezpieczenia emerytalne i rentowe (numeryczna wartość zmiennej *ubezp_podst_skladki_uber_ZUS*). Rekord, dla którego podstawa ta była maksymalna, uznany został za rekord dotyczący pracy głównej. W przypadku występowania więcej niż jednego takiego rekordu, rekordy z podstawą składki niższą niż maksymalna należało uznać za prace dodatkowe, po czym przejść do wyboru pracy głównej na bazie hierarchii kodów tytułu ubezpieczenia (zmienna *ubezp_kod_titul_ubez_ZUS*). Hierarchia kodów tytułu ubezpieczenia jest przedstawiona w tablicy nr 1.

Tablica 1. Hierarchia kodów tytułu ubezpieczenia do wyznaczenia pracy głównej

Nr w hierarchii	Kod tytułu ubezpieczenia
1	0510

Nr w hierarchii	Kod tytułu ubezpieczenia
2	0512
3	0513
4	0544
5	0545
6	0570
7	0572
8	0520
9	0530
10	0540
11	0511
12	0543
13	0580
14	0581
15	0610
16	2010
17	2020
18	1130
19	1131
20	1132
21	1133
22	1134
23	1135
24	1140
25	1120
26	1110
27	1111
28	1112
29	1113
30	1114
31	1010
32	1011
33	1012
34	0110
35	0111
36	0200
37	0310
38	0320
39	0800
40	1240
41	1211

W pierwszej kolejności należało odnaleźć rekordy z kodem tytułu ubezpieczenia położonym najwyżej w hierarchii, czyli „0510”. Jeśli występował tylko jeden taki rekord, uznano, że opisuje on pracę główną. Jeśli nie występował żaden rekord z kodem „0510”, należało w analogiczny sposób sprawdzić kolejny kod w przedstawionej powyżej hierarchii kodów tytułu ubezpieczenia. Jeżeli natomiast dany kod pojawił się u danej osoby więcej niż raz, wówczas rekordy z innymi kodami tytułu ubezpieczenia należało uznać za prace dodatkowe i sprawdzić datę początku obowiązywania składek ZUS (zmienna *ubezp_data_skladki_ZUS*). Za pracę główną należało uznać tę, dla której data początku obowiązywania składek była najwcześniejsza. W przypadku wystąpienia więcej niż jednej takiej pracy, należało przyjąć za pracę główną tę, która w rejestrze ZUS występowała z najniższym numerem rekordu (zmienna *LP*).

Po przeprowadzeniu procedury wyboru pracy głównej otrzymano zbiór pracujących w pracy głównej wg ZUS, w którym każda osoba ubezpieczona z tytułu świadczenia pracy występowała tylko raz.

Przeprowadzenie powyższej procedury zakończyło proces wyboru pracy głównej wg ZUS. Nie było konieczności wdrożenia dalszych działań.

- 22) zidentyfikowanie osób, które w okresie referencyjnym posiadały status zatrudnionego wg AZD.

Do osób zatrudnionych zaliczono osoby spełniające poniższe warunki:

- a) byli w wieku 16 lat i więcej,
- b) 4 pierwsze znaki kodu tytułu ubezpieczenia były równe „0110”, „0111”, „0610”, „0800”, „1133”, „1134”, „1140”, „2010”, „2020”.

Wśród osób zatrudnionych według AZD znalazły się osoby objęte odrębnym systemem zaopatrzenia emerytalnego, dla których w zbiorze ZUS nie wystąpiła informacja o wysokości podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe. Należały do nich osoby ubezpieczone z kodami tytułu ubezpieczenia: „1133”, „1140”, „2010”, „2020”. W związku z tym zbiorowość zatrudnionych została ograniczona do osób ubezpieczonych w okresie referencyjnym z kodami tytułu ubezpieczenia „0110”, „0111”, „0610”, „0800”, „1134”.

- 23) ustalenie podstawowego/nominalnego wymiaru etatu dla osób zatrudnionych.

Utworzono zmienną *wymiar_etatu_nominalny* opisującą nominalny wymiar etatu w następujący sposób:

- a) w przypadku osób ubezpieczonych z kodami tytułu ubezpieczenia „0610”, „1134” przyjęto, że były one zatrudnione w pełnym nominalnym wymiarze etatu (zmienna *wymiar_etatu_nominalny* przyjęła wartość „1,000”);
- b) w przypadku osób ubezpieczonych z kodami tytułu ubezpieczenia „0110”, „0111”, „0800” przyjęto, że za wartość zmiennej *wymiar_etatu_nominalny* dla tych osób należy uznać:
 - wartość „1,000”, jeżeli licznik i mianownik wymiaru czasu pracy w ZUS były sobie równe, niezależnie od przyjmowanych wartości,
 - wartość ilorazu licznika i mianownika zaokrąglona do trzech miejsc po przecinku (o ile $\text{licznik} > 0$, $\text{mianownik} > 0$, $\text{licznik} \leq \text{mianownik}$ i $\text{licznik}/\text{mianownik} \geq 0,001$);
- c) w pozostałych przypadkach przyjęto, że wartość zmiennej *wymiar_etatu_nominalny* pozostała nieustalona.

Dla osób z nieustalonym nominalnym wymiarem etatu nie było możliwe wykonanie dalszych prac mających na celu naliczenie miar wynagrodzeń. Kolejne etapy przeprowadzono tylko dla tych obserwacji ze zbiorowości zatrudnionych, dla których nominalny wymiar etatu przyjął wartość większą lub równą „0,001” i jednocześnie mniejszą lub równą „1,000”.

- 24) ustalenie rzeczywistego wymiaru etatu w okresie referencyjnym dla osób zatrudnionych

Pod uwagę należało wziąć okresy przerw w opłacaniu składek, okresy pobierania świadczeń, a także inne czynniki mające wpływ na podstawę wymiaru składek emerytalno-rentowych. Do ustalenia rzeczywistego wymiaru etatu wykorzystano następujące zmienne ze zbioru ZUS:

- a) *ubezyp_data_skladki_ZUS* – data początku obowiązywania składek ZUS;
- b) *ubezyp_kod_sw_ZUS* – kod świadczenia lub przerwy w płaceniu składek ZUS¹;
- c) *ubezyp_data_od_sw_ZUS* – data początku okresu za jaki nastąpiła wypłata świadczenia lub przerwa w opłacaniu składki ZUS;
- d) *ubezyp_data_do_sw_ZUS* – data końca okresu jaki nastąpiła wypłata świadczenia lub przerwa w opłacaniu składki ZUS;

¹wykaz kodów zawiera załącznik nr 2.

e) *ubezp_kod_przechr_podst_uber_ZUS* – kod informacji o przekroczeniu rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe.

Do zbioru ZUS dodano poniższe zmienne pomocnicze:

a) *dni_okres_ref* – liczba dni kalendarzowych w miesiącu referencyjnym;

b) *dni_okres_ref_poprz* – liczba dni kalendarzowych w miesiącu poprzedzającym miesiąc referencyjny;

c) *jest_przerwa* – informacja o wystąpieniu kodu świadczenia/przerwy obniżającego rzeczywisty wymiar etatu (zmienna przyjęła wartość „1”, jeśli wystąpił kod świadczenia/przerwy o symbolu „1xx”, wartość „3”, jeśli wystąpił kod świadczenia/przerwy o symbolu „3xx” i wartość „0”, jeśli przerwa nie wystąpiła);

d) *mc_przerwy* – informacja, czy przerwa wystąpiła w miesiącu referencyjnym (zmienna przyjęła wartość „1”), w miesiącu poprzedzającym miesiąc referencyjny (zmienna przyjęła wartość „2”) czy w innym miesiącu lub jeśli nie wystąpiła (zmienna przyjęła wartość nieustaloną);

e) *przerwa_caly_mc* – informacja, czy przerwa objęła cały miesiąc referencyjny lub miesiąc poprzedzający miesiąc referencyjny (zmienna przyjęła wartość „1”, gdy przerwa objęła cały miesiąc referencyjny lub cały miesiąc poprzedzający miesiąc referencyjny i wartość „0” w pozostałych przypadkach);

f) *przerwa_rok_ref* – informacja, czy przerwa dla kodów świadczenia/przerwy o symbolach „3xx” wystąpiła w roku referencyjnym (zmienna przyjęła wartość „1”, gdy przerwa w całości zawierała się w roku referencyjnym, wartość „2”, gdy przerwa rozpoczęła się przed początkiem roku referencyjnego, a zakończyła się w roku referencyjnym, wartość „0”, gdy przerwa w całości odbyła się przed rokiem referencyjnym; w pozostałych przypadkach wartość zmiennej pozostała nieustaloną);

g) *liczba_dni_przerwy* – informacja, ile dni przerwy z kodem o symbolu „3xx” wystąpiło w roku referencyjnym (w przypadku, gdy przerwa nie występowała lub w całości wystąpiła przed rozpoczęciem roku referencyjnego, zmienna przyjęła wartość „0”, a w przypadku, gdy wystąpiła przerwa z kodem o symbolu „1xx”, to zmienna pozostała nieustaloną);

h) *liczba_dni_bez_skladki* – informacja o liczbie dni, które wystąpiły pomiędzy początkiem miesiąca referencyjnego, a datą początku obowiązywania składek, jeżeli data początku obowiązywania składek wystąpiła w miesiącu referencyjnym (liczone od pierwszego dnia miesiąca do dnia poprzedzającego datę początku obowiązywania składek); jeżeli data początku obowiązywania składek wystąpiła najpóźniej w pierwszym dniu kalendarzowym miesiąca referencyjnego, to zmienna przyjęła wartość nieustaloną.

Dla każdej obserwacji ustalono liczbę dni oskładkowanych mających wpływ na rzeczywisty wymiar etatu. Posłużyła do tego zmienna *liczba_dni_skladki*, której wartości ustalono następująco:

a) jeżeli w okresie/miesiącu referencyjnym nie wystąpił kod świadczenia/przerwy obniżający rzeczywisty wymiar etatu (zmienna *jest_przerwa* przyjęła wartość „0”), to:

– jeżeli data początku obowiązywania składek była wcześniejsza niż początek miesiąca poprzedzającego miesiąc referencyjny, to przyjęto, że cały miesiąc referencyjny był oskładkowany i liczba dni składki była równa liczbie dni miesiąca referencyjnego

$liczba_dni_skladki = dni_okres_ref$,

– jeżeli data początku obowiązywania składek przypadła w miesiącu referencyjnym, to w liczbie dni oskładkowanych należało uwzględnić datę początku obowiązywania składek

$liczba_dni_skladki = dni_okres_ref - liczba_dni_bez_skladki$,

– jeżeli data początku obowiązywania składek przypadła w miesiącu poprzedzającym miesiąc referencyjny, to nie można było ustalić liczby dni oskładkowanych ze względu na brak informacji, czy składki dotyczyły miesiąca referencyjnego, czy też miesiąca poprzedzającego miesiąc referencyjny i zmienna *liczba_dni_skladki* przyjęła wartość nieustaloną;

b) jeżeli w okresie/miesiącu referencyjnym wystąpił kod świadczenia/przerwy obniżający rzeczywisty wymiar etatu i miał on symbol „1xx” (zmienna *jest_przerwa* przyjęła wartość „1”), to nie

było możliwości ustalenia, czy przerwa ta została ujęta w podstawie wymiaru składek na ubezpieczenia emerytalne i rentowe i zmienna *liczba_dni_skladki* przyjęła wartość nieustaloną;

c) jeżeli w okresie/miesiącu referencyjnym wystąpił kod świadczenia/przerwy obniżający rzeczywisty wymiar etatu i miał on symbol „3xx” (zmienna *jest_przerwa* przyjęła wartość „3”) i data początku obowiązywania składek była wcześniejsza niż początek miesiąca poprzedzającego miesiąc referencyjny, to:

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była równa lub większa od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy \geq dni_okres_ref$), to przyjęto, że cały miesiąc referencyjny nie był oskładkowany i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy < dni_okres_ref$), ale okres świadczenia/przerwy obejmował cały miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „1”) lub miesiąc poprzedzający miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „2”), to przyjęto, że nie wystąpiły dni oskładkowane i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy < dni_okres_ref$) i okres świadczenia/przerwy nie obejmował całego miesiąca referencyjnego ani miesiąca poprzedzającego miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość różną od „1” i różną od „2”), to liczbę dni oskładkowanych obniżono o długość okresu świadczenia/przerwy

$liczba_dni_skladki = dni_okres_ref - liczba_dni_przerwy$;

d) jeżeli w okresie/miesiącu referencyjnym wystąpił kod świadczenia/przerwy obniżający rzeczywisty wymiar etatu i miał on symbol „3xx” (zmienna *jest_przerwa* przyjęła wartość „3”) i data początku obowiązywania składek przypadła w miesiącu referencyjnym, to:

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była równa lub większa od liczby dni w miesiącu referencyjnym, przyjęto, że cały miesiąc referencyjny nie był oskładkowany i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym, ale okres świadczenia/przerwy obejmował cały miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „1”) lub miesiąc poprzedzający miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „2”), przyjęto, że nie wystąpiły dni oskładkowane i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy < dni_okres_ref$) i okres świadczenia/przerwy nie obejmował całego miesiąca referencyjnego ani miesiąca poprzedzającego miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość różną od „1” i różną od „2”), to liczbę dni oskładkowanych obniżono o liczbę dni, które wystąpiły pomiędzy początkiem miesiąca referencyjnego, a datą początku obowiązywania składek oraz o długość okresu świadczenia/przerwy

$liczba_dni_skladki = (dni_okres_ref - liczba_dni_bez_skladki) - liczba_dni_przerwy$;

e) jeżeli w okresie/miesiącu referencyjnym wystąpił kod świadczenia/przerwy obniżający rzeczywisty wymiar etatu i miał on symbol „3xx” (zmienna *jest_przerwa* przyjęła wartość „3”) i data początku obowiązywania składek przypadła w miesiącu poprzedzającym miesiąc referencyjny, to:

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była równa lub większa od liczby dni w miesiącu referencyjnym, przyjęto, że cały miesiąc referencyjny nie był oskładkowany i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym, ale okres świadczenia/przerwy obejmował cały miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „1”) lub miesiąc poprzedzający miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość „2”), przyjęto, że nie wystąpiły dni oskładkowane i zmienna *liczba_dni_skladki* przyjęła wartość „0”,

- jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy < dni_okres_ref$) i okres

świadczenia/przerwy nie obejmował całego miesiąca referencyjnego ani miesiąca poprzedzającego miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość różną od „1” i różną od „2”) oraz okres świadczenia/przerwy zakończył się w miesiącu referencyjnym, to liczbę dni oskładkowanych obniżono o długość okresu świadczenia/przerwy $liczba_dni_skladki = dni_okres_ref - liczba_dni_przerwy$,

– jeżeli liczba dni świadczenia/przerwy obejmujących rok referencyjny była mniejsza od liczby dni w miesiącu referencyjnym ($liczba_dni_przerwy < dni_okres_ref$) i okres świadczenia/przerwy nie obejmował całego miesiąca referencyjnego ani miesiąca poprzedzającego miesiąc referencyjny (zmienna *mc_przerwy* przyjęła wartość różną od „1” i różną od „2”) oraz okres świadczenia/przerwy zakończył się w miesiącu poprzedzającym miesiąc referencyjny, to nie można było ustalić liczby dni oskładkowanych i zmienna *liczba_dni_skladki* przyjęła wartość nieustaloną.

Utworzono zmienną *wymiar_etatu_rzeczywisty* opisującą rzeczywisty wymiar czasu pracy w okresie referencyjnym, której wartość została zdefiniowana wzorem:

$$\frac{liczba_dni_skladki}{dni_okres_ref} * wymiar_etatu_nominalny$$

Wartość zmiennej była ustalana tylko dla rekordów, dla których wartości zmiennych składowych były niepuste i wyliczana z dokładnością do trzech miejsc po przecinku.

Dalsze prace prowadzące do naliczenia miar wynagrodzeń wykonano tylko dla tych obserwacji ze zbiorowości zatrudnionych, dla których rzeczywisty wymiar etatu przyjął ustaloną wartość większą lub równą „0.001” i jednocześnie mniejszą lub równą „1.000”.

25) oszacowanie wynagrodzenia brutto przypadającego na pełen etat – przeliczenie rzeczywistego wynagrodzenia brutto wypłaconego osobie zatrudnionej w okresie referencyjnym w odniesieniu do rzeczywistego wymiaru etatu w okresie referencyjnym;

Dodano dwie zmienne niezbędne do oszacowania wynagrodzeń brutto dla poszczególnych obserwacji, bazujące na źródłowej zmiennej *ubezp_podst_skladki_uber_ZUS*:

a) zmienną *wyn_rzeczywiste* zawierającą informację o podstawie wymiaru składki emerytalno-rentowej uznawanej na potrzeby niniejszej pracy badawczej za wynagrodzenie rzeczywiste, czyli nieprzeliczone do pełnego etatu;

b) zmienną *wyn_na_etat* zawierającą informację o wynagrodzeniu rzeczywistym (podstawie wymiaru składki emerytalno-rentowej) przeliczonym do pełnego etatu zgodnie z poniższym wzorem:

$$wyn_na_etat = \frac{wyn_rzeczywiste}{wymiar_etatu_rzeczywisty}$$

26) uzupełnienie informacji o wynagrodzeniach brutto dla osób, które przekroczyły roczną podstawę wymiaru składek na ubezpieczenia emerytalne i rentowe (jedynie dla zbiorów, dla których istniała możliwość porównania ze zbiorem za wcześniejszy okres referencyjny);

Dla zatrudnionych, którzy w miesiącu referencyjnym przekroczyli roczną podstawę wymiaru składek na ubezpieczenia emerytalne i rentowe wartości zmiennych *wyn_rzeczywiste* i *wyn_na_etat* nie były adekwatne do wynagrodzeń brutto wypłaconych w miesiącu referencyjnym, a dla osób, które przekroczyły roczną podstawę wymiaru składek na ubezpieczenia emerytalne i rentowe w miesiącach wcześniejszych obie zmienne przyjmowały wartość „0”.

W okresach referencyjnych: grudzień 2016 r. oraz czerwiec 2017 r., wartości zmiennych *wymiar_etatu_rzeczywisty*, *wyn_rzeczywiste* i *wyn_na_etat* dla osób, które w miesiącu referencyjnym przekroczyły roczną podstawę wymiaru składek na ubezpieczenia emerytalne i rentowe, zostały częściowo imputowane wartościami zmiennych *wymiar_etatu_rzeczywisty*, *wyn_rzeczywiste* i *wyn_na_etat* z wcześniejszych okresów referencyjnych (z września 2016 r. dla grudnia 2016 r. i z marca 2017 r. dla czerwca 2017 r.) z danego roku referencyjnego. Było to możliwe przy zachowaniu warunków:

a) wartości zmiennych w obu okresach referencyjnych odnosiły się do tego samego zatrudnionego (te same wartości zmiennej *ident_osoby* w obu zbiorach), był on ubezpieczony u tego

samemu płatnika (te same wartości zmiennej *platnik_NIP_ZUS* lub zmiennej *platnik_REGON_ZUS* w obu zbiorach) z tym samym kodem tytułu ubezpieczenia (te same wartości zmiennej *ubezp_kod_tytuł_ubez_ZUS* w obu zbiorach) i tą samą datą początku obowiązywania składek (te same wartości zmiennej *data_początku_składki* w obu zbiorach) oraz tym samym nominalnym wymiarem etatu (te same wartości zmiennej *wymiar_etatu_nominalny* w obu zbiorach);

b) kombinacja wartości zmiennych wyszczególnionych w punkcie a) była unikalna w zbiorze za wcześniejszy okres referencyjny (jeśli dana kombinacja wartości zmiennych wystąpiła w więcej niż jednym rekordzie zbioru za wcześniejszy okres referencyjny, to wartości zmiennych *wymiar_etatu_rzeczywisty*, *wyn_rzeczywiste* i *wyn_na_etat* nie były imputowane);

c) w zbiorze za wcześniejszy okres referencyjny wartość zmiennej *wyn_rzeczywiste* była większa od zera oraz kod informacji o przekroczeniu rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe był niewypełniony.

Powyższego uzupełnienia dokonywano tylko w ramach tego samego roku referencyjnego. Dla okresów referencyjnych czerwiec i wrzesień 2016 r. nie było możliwości uzupełnienia wartości zmiennych *wymiar_etatu_rzeczywisty*, *wyn_rzeczywiste* i *wyn_na_etat* informacjami z innych okresów referencyjnych ze względu na brak unikalnego identyfikatora osoby (brak numeru PESEL) oraz brak informacji o przekroczeniu rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe w zbiorach za marzec i czerwiec 2016 roku.

27) wybór rekordów, dla których wynagrodzenie rzeczywiste miało ustaloną wartość większą od zera przy czym:

a) dla zbiorów za wrzesień 2016 r. i marzec 2017 r. nie wystąpiła informacja o przekroczeniu rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe (zmienna *ubezp_kod_przekr_podst_uber_ZUS* była niewypełniona);

b) dla zbiorów za grudzień 2016 r. i czerwiec 2017 r. nie wystąpiła informacja o przekroczeniu rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe (zmienna *ubezp_kod_przekr_podst_uber_ZUS* była niewypełniona) lub jeśli wystąpiła, to wartości zmiennych *wymiar_etatu_rzeczywisty*, *wyn_rzeczywiste* i *wyn_na_etat* zostały zaimputowane informacjami z wcześniejszych okresów referencyjnych.

Przeprowadzenie powyższej procedury zakończyło proces uzupełniania informacji o wynagrodzeniach oraz wyboru rekordów, dla których wynagrodzenia były możliwe do wyliczenia.

28) ustalenie wartości dla wszystkich zmiennych niezbędnych do naliczenia wskaźników i danych wynikowych określonych w celu pracy badawczej.

Do naliczenia wskaźników wynikowych należało przygotować następujące zmienne:

a) identyfikacyjne:

- *LP_ZUS* – zmienna zawierająca informację o numerze rekordu w zbiorze ZUS,
- *ident_osoby* – zmienna zawierająca informację o identyfikatorze osoby;

b) osobowe:

- *ubezp_plec_uzup* – zmienna zawierająca informację o płci osoby,
- *ubezp_wiek_ZUS* – zmienna zawierająca informację o wieku osoby,
- *sekcja_PKD* – zmienna zawierająca informację o sekcji PKD 2007 płatnika osoby,
- *wymiar_czasu_pracy_ZUS* – zmienna zawierająca informację o wymiarze czasu pracy osoby;

c) adresowe:

- *ubezp_adr_zam_kod_gmn_ZUS* – zmienna zawierająca informację o kodzie gminy adresu zamieszkania osoby,
- *platnik_adr_s_kod_gmn_ZUS* – zmienna zawierająca informację o kodzie gminy adresu miejsca siedziby pracy głównej osoby.

Wśród wartości wszystkich wymienionych wyżej zmiennych nie wystąpiły braki danych.

Powyższe czynności zostały wykonane również na zbiorach za marzec i czerwiec 2016 r. Zastosowano jednak pewne modyfikacje spowodowane brakiem części zmiennych dla tych okresów. Ze względu na brak numeru PESEL ubezpieczonego, zrezygnowano z bazującego na numerze PESEL uzupełniania zmiennych opisujących płeć, wiek i adres zamieszkania osoby ubezpieczonej. Braki wartości zmiennych dotyczących płci i wieku uzupełniono przy zastosowaniu metody imputacji wielokrotnej na poziomie danych jednostkowych opisanej w podrozdziale 2.3.8.

Ze względu na ograniczoną liczbę zmiennych adresowych oraz uwzględniając cel pracy badawczej, na potrzeby wynagrodzeń zdecydowano dla zbiorów za marzec i czerwiec 2016 r. ustalać adres zamieszkania osoby ubezpieczonej z dokładnością do kodu TERYT powiatu. Ustalenie kodu TERYT powiatu zamieszkania osoby ubezpieczonej zostało wykonane w następujących krokach:

1) uzupełnienie zmiennych opisujących adres zamieszkania ubezpieczonego wartościami zmiennych opisujących adres zameldowania ubezpieczonego (zgodnie z zasadami przyjętymi w ZUS, adres zamieszkania jest wypełniony jedynie wtedy, gdy jest on różny od adresu zameldowania), o ile wszystkie zmienne dotyczące adresu zamieszkania ubezpieczonego były niewypełnione;

2) uzupełnienie zmiennej opisującej kod pocztowy adresu zamieszkania wartością zmiennej opisującej kod pocztowy adresu zameldowania, w przypadku gdy wartości zmiennych opisujących nazwę gminy oraz nazwę miejscowości są takie same dla obu adresów;

3) uzupełnienie zmiennej opisującej nazwę gminy adresu zamieszkania wartością zmiennej opisującej nazwę gminy adresu zameldowania, w przypadku gdy wartości zmiennych opisujących kod pocztowy oraz nazwę miejscowości są takie same dla obu adresów;

4) połączenie zbioru ZUS ze zbiorem PNA oraz uzupełnienie informacji o kodzie TERYT powiatu adresu zamieszkania ubezpieczonego w oparciu o zmienne pochodzące ze zbioru PNA; utworzenie zmiennej *ubezp_adr_zam_kod_pow* opisującej kod TERYT powiatu adresu zamieszkania ubezpieczonego i uzupełnienie jej kodem TERYT powiatu ze zbioru PNA w następujących krokach:

a) dla rekordów ze zbioru ZUS mających wypełnioną zmienną opisującą kod pocztowy adresu zamieszkania, dla których w zbiorze PNA kod ten występuje w tylko jednym powiecie, za wartość zmiennej przyjęto kod TERYT tego powiatu,

b) dla rekordów ze zbioru ZUS mających wypełnioną zmienną opisującą kod pocztowy adresu zamieszkania oraz zmienną opisującą nazwę miejscowości adresu zamieszkania, dla których w zbiorze PNA powiązanie tego kodu pocztowego z tą nazwą miejscowości występuje w tylko jednym powiecie, za wartość zmiennej przyjęto kod TERYT tego powiatu,

c) dla rekordów ze zbioru ZUS mających wypełnioną zmienną opisującą kod pocztowy adresu zamieszkania oraz zmienną opisującą nazwę gminy adresu zamieszkania, dla których w zbiorze PNA powiązanie tego kodu pocztowego z tą nazwą gminy występuje w tylko jednym powiecie, za wartość zmiennej przyjęto kod TERYT tego powiatu,

d) dla rekordów ze zbioru ZUS mających wypełnioną zmienną opisującą nazwę gminy adresu zamieszkania oraz zmienną opisującą nazwę miejscowości adresu zamieszkania, dla których w zbiorze PNA powiązanie tej nazwy gminy z tą nazwą miejscowości występuje w tylko jednym powiecie, za wartość zmiennej przyjęto kod TERYT tego powiatu;

5) połączenie zbioru ZUS ze zbiorem TERYT za 2016 rok oraz weryfikacja zgodności kodów TERYT powiatu adresu zamieszkania ubezpieczonego występujących w zbiorze ZUS z kodami TERYT powiatu występującymi w zbiorze TERYT za 2016 rok;

6) pozostałe braki wartości zmiennej *ubezp_adr_zam_kod_pow* uzupełniono metodą imputacji na podstawie rozkładu prawdopodobieństwa w połączeniu z imputacją wielokrotną; proces imputacji został opisany w podrozdziale 2.3.8.

Ze względu na brak numeru PESEL ubezpieczonego w tych okresach niemożliwe było połączenie zbioru ZUS ze zbiorem KEP, a tym samym uzyskanie informacji o osobach zamieszkujących poza granicami kraju. Z tego powodu wszystkie osoby zostały uznane za mieszkające na terytorium Polski i wszystkim został przypisany kod TERYT powiatu zamieszkania.

W zbiorach za marzec i czerwiec 2016 roku mogły pozostać obserwacje, dla których nastąpiło przekroczenie rocznej podstawy wymiaru składek na ubezpieczenia emerytalne i rentowe, ale ze względu na brak zmiennej *ubezp_kod_przekr_podst_uber_ZUS* nie było możliwości wyodrębnienia tych obserwacji.

W zbiorach za marzec 2016 r. i czerwiec 2016 r. nie wystąpiły również zmienne informujące o okresie, za jaki nastąpiła wypłata świadczenia lub przerwa w opłacaniu składki (zmienne *ubezp_data_od_sw_ZUS* i *ubezp_data_od_sw_ZUS*). Nie można było zatem ustalić, czy świadczenie/przerwa miało wpływ na wysokość podstawy składki emerytalno-rentowej. Do zbiorów za marzec i czerwiec 2016 r. nie dodano zmiennych *mc_przerwy*, *przerwa_caly_mc*, *przerwa_rok_ref* i *liczba_dni_przerwy*. Przy ustalaniu wartości zmiennej *liczba_dni_skladki* dla poszczególnych obserwacji pod uwagę wzięto tylko zmienne *jest_przerwa*, *dni_okres_ref* i *liczba_dni_bez_skladki*, a dla rekordów, dla których zmienna *jest_przerwa* miała wartość „3” przyjęto, że zmienna *liczba_dni_skladki* ma wartość nieustaloną.

Pozostałe czynności na zbiorach ZUS na zbiorach za marzec i czerwiec 2016 r. przeprowadzono bez zmian w stosunku do pozostałych badanych okresów.