Harmonisation Summary Report Template

Table of contents

# List of Tables

[Table S1](#supptbl-cohort-introduction): Cohorts used in study.

[Table S2](#supptbl-age-harmonisation): Harmonisation process of *age*.

[Table S3](#supptbl-sex-harmonise-values): Harmonised values of *sex*.

[Table S4](#supptbl-sex-harmonisation): Harmonised process of *sex*.

[Table S5](#X4398fa6d79f36859531308484c96599922afe06): Harmonisation process of *height\_cm*, *weight\_kg*, *bsa\_m2* and *bmi*.

[Table S6](#supptbl-smoking-history-harmonise-values): Harmonised values of *smoke\_current* and *smoke\_past*.

[Table S7](#supptbl-smoking-history-harmonisation): Harmonised process of *smoke\_current* and *smoke\_past*.

[Table S8](#supptbl-smoking-history-validation): Valid values of *smoke\_current* and *smoke\_past*.

[Table S9](#supptbl-have-sob-harmonise-values): Harmonised values of *have\_sob*.

[Table S10](#supptbl-have-sob-harmonisation): Harmonised process of *have\_sob*.

[Table S11](#supptbl-have-chest-pain-harmonise-values): Harmonised values of *have\_chest\_pain*.

[Table S12](#supptbl-have-chest-pain-harmonisation): Harmonised process of *have\_chest\_pain*.

[Table S13](#supptbl-symptoms-harmonise-values): Harmonised values of *symptoms*.

[Table S14](#supptbl-symptoms-harmonisation): Harmonised process of *symptoms*.

[Table S15](#Xc3f4b168d9323bff98acb411cedd9331ed8d99b): Harmonised values of *chest\_pain\_type*.

[Table S16](#X39d28d39dfb0faced0559023910e984af15abf6): Harmonised process of *chest\_pain\_type*.

[Table S17](#supptbl-chest-pain-character-validation): Valid values of *symptoms* and *chest\_pain\_type*.

# 1. Cohort

## 1.1 Cohort Introduction

Here are the cohorts used in the paper.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table S1: Cohorts used in study.   | Cohort ID | Cohort Name | | --- | --- | | Cohort A | Cohort A Full Name | | Cohort B | Cohort B Full Name | |

# 2. Demographic and Clinical Variables

Steps taken to harmonised data columns in relation to demographic and clinical variables are discussed here.

## 2.1 Age

*age\_years* is the harmonised positive integer data field to denote the age of the patient during the time of the CT scan.

It is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S2: Harmonisation process of *age\_years*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *age* of positve integer values. 0 is used to indicate unknown values. | Value of *0* in *age* will be changed to *NA*. *age\_years* will take the values of *age*. | | Cohort B | Column *Age* of positve integer values | *age\_years* will take the values of *Age*. | |

## 2.2 Sex

*sex* is the harmonised data field to denote the sex of the patient during the time of the CT scan.

It holds the following values:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S3: Harmonised values of *sex*.   | Value | Description | | --- | --- | | 0 | female | | 1 | male | | -1 | unknown | |

It is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S4: Harmonisation process of *sex*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *sex* with  *F* as female.  *M* as male. | Change the values of *sex* as follows:  *F* as 0.  *M* as 1. | | Cohort B | Column *Sex* with  *Female* as female.  *Male* as male. | Map the values of *Sex* to *sex* as follows:  *Female* as 0.  *Male* as 1. | |

## 2.3 Height, Weight, BMI and BSA

*height* is the harmonised positive real data field to denote the height in cm of the patient during the time of the CT scan.

*weight* is the harmonised positive real data field to denote the weight in kg of the patient during the time of the CT scan.

*bsa\_m2* is the harmonised positive real data field to denote the body surface area in m2 of the patient during the time of the CT scan.

*bmi* is the harmonised positive real data field to denote the body mass index of the patient during the time of the CT scan.

All values are converted to two decimal places if the number of decimal places exceeded two.

They are harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S5: Harmonisation process of *height\_cm*, *weight\_kg*, *bsa\_m2* and *bmi*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *height* in cm of positve real numeric values in one decimal place.  Column *weight* in kg of positve real numeric values in one decimal place. | *height\_cm* will take the values of *height*.  *weight\_kg* will take the values of *weight*.  *bsa\_m2* and *bmi* are calculated using data fields *height\_cm* and *weight\_kg*. All values are then converted to two decimal places. | | Cohort B | Column *Height* in cm of positve integer values.  Column *Weight* in kg of positve integer values. | *height\_cm* will take the values of *Height*.  *weight\_kg* will take the values of *Weight*.  *bsa\_m2* and *bmi* are calculated using data fields *height\_cm* and *weight\_kg*. All values are then converted to two decimal places. | |

## 2.4 Smoking History

*smoke\_current* is the harmonised data field to denote if the patient is a current smoker during the time of the CT scan. *smoke\_past* is the harmonised data field to denote if the patient is a past smoker during the time of the CT scan.

They hold the following values:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S6: Harmonised values of *smoke\_current* and *smoke\_past*.   | Value | Description | | --- | --- | | 0 | no | | 1 | yes | | -1 | unknown | |

They are harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S7: Harmonised process of *smoke\_current* and *smoke\_past*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *smoke\_current\_good* with  *0* as no.  *1* as yes.  *-1* as unknown.  Column *smoke\_past\_good* with  *0* as no.  *1* as yes.  *-1* as unknown. | *smoke\_current* will take the values of *smoke\_current\_good*.  *smoke\_past* will take the values of *smoke\_past\_good*. | | Cohort B | Column *Smoke History* with  *non-smoker* as non-smoker.  *past smoker* as a past smoker.  *current smoker* as a current smoker.  *NA* as unknown. | Map the values of *Smoke History* to *smoke\_current* as follows:  *non-smoker* and *past smoker* as 0.  *current smoker* as 1.  *NA* as -1.  Map the values of *Smoke History* to *smoke\_past* as follows:  *non-smoker* and *current smoker* as 0.  *past smoker* as 1.  *NA* as -1. | |

After harmonisation, we validate the values of *smoke\_current* and *smoke\_past* to ensure that there can only be the following cases:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S8: Valid values of *smoke\_current* and *smoke\_past*.   | Description | *smoke\_current* | *smoke\_past* | | --- | --- | --- | | Non-smoker | 0 | 0 | | Past smoker | 0 | 1 | | Current smoker | 1 | 0 | | Unknown | -1 | -1 | |

## 2.5 Have Shortness of Breath

*have\_sob* is the harmonised data field to denote if the patient has shortness of breath during the time of the CT scan.

It holds the following values:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S9: Harmonised values of *have\_sob*.   | Value | Description | | --- | --- | | 0 | no | | 1 | yes | | -1 | unknown | |

*have\_sob* is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S10: Harmonised process of *have\_sob*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *have\_sob* with  *0* as no.  *1* as yes. | *have\_sob* remains unchanged. | | Cohort B | Column *Dyspnea* with  *no* as no.  *yes* as yes. | Map the values of *Dyspnea* to *have\_sob* as follows:  *no* as 0.  *yes* as 1. | |

## 2.6 Have Chest Pain

*have\_chest\_pain* is the harmonised data field to denote if the patient has chest pain during the time of the CT scan.

It holds the following values:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S11: Harmonised values of *have\_chest\_pain*.   | Value | Description | | --- | --- | | 0 | no | | 1 | yes | | -1 | unknown | |

*have\_chest\_pain* is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S12: Harmonised process of *have\_chest\_pain*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *chest\_pain\_type* with  *0* as no chest pain.  *1* as typical chest pain.  *2* as atypcial chest pain.  *3* as nonanginal chest pain. | Map the values of *chest\_pain\_type* to *have\_chest\_pain* as follows:  If *chest\_pain\_type* has a value of *1*, *2* or *3*, *have\_chest\_pain* will be 1.  Else, if *chest\_pain\_type* has a value of *0*, *have\_chest\_pain* will be 0. | | Cohort B | Column *Chest Pain Character* with  *no chest pain* as no chest pain.  *typical* as typical chest pain.  *atypcial* as atypcial chest pain.  *nonanginal* as nonanginal chest pain. | Map the values of *Chest Pain Character* to *have\_chest\_pain* as follows:  If *chest\_pain\_type* has a value of *typical*, *atypical* or *nonanginal*, *have\_chest\_pain* will be 1.  Else, if *chest\_pain\_type* has a value of *no chest pain*, *have\_chest\_pain* will be 0. | |

## 2.7 Symptoms

*symptoms* is the harmonised data field to denote the patient’s symptoms during the time of the CT scan.

It holds the following values:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S13: Harmonised values of *symptoms*.   | Value | Description | | --- | --- | | 0 | asymptomatic | | 1 | chest pain | | 2 | only dyspnea | | 3 | others | | -1 | unknown | |

Regarding the symptoms: chest pain, dypsnea and other symptoms:

* If a patient has all three symptoms, chest pain will take the highest priority. Hence, *symptoms* = 1
* If a patient has both dyspnea and other symptoms (not chest pain related), dyspnea will take the higher priority. Hence, *symptoms* = 2

The general approach is to assume that the patients are asymptomatic (*symptoms* = 0) unless indicated that they have chest pain (*symptoms* = 1), dypsnea (*symptoms* = 2), other symptoms like heart palpitations (*symptoms* = 3) or all symptom related data fields are missing (*symptoms* = -1).

*symptoms* is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S14: Harmonised process of *symptoms*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *have\_sob* with  *0* as no.  *1* as yes.  Column *chest\_pain\_type* with  *0* as no chest pain.  *1* as typical chest pain.  *2* as atypcial chest pain.  *3* as nonanginal chest pain. | Map the values of *chest\_pain\_type* and *have\_sob* to *symptoms* as follows:  If *chest\_pain\_type* has a value of *1*, *2* or *3*, *symptoms* will be 1.  Else, if *chest\_pain\_type* has a value of *0* and *have\_sob* has a value of *1*, *symptoms* will be 2.  Else, *symptoms* will be 0. | | Cohort B | Column *Dyspnea* with  *no* as no.  *yes* as yes.  Column *Chest Pain Character* with  *no chest pain* as no chest pain.  *typical* as typical chest pain.  *atypcial* as atypcial chest pain.  *nonanginal* as nonanginal chest pain. | Map the values of *Chest Pain Character* and *Dyspnea* to *symptoms* as follows:  If *chest\_pain\_type* has a value of *typical*, *atypical* or *nonanginal*, *symptoms* will be 1.  Else, if *chest\_pain\_type* has a value of *no chest pain* and *Dyspnea* has a value of *yes*, *symptoms* will be 2.  Else, *symptoms* will be 0. | |

## 2.8 Chest Pain Type

*chest\_pain\_type* is the harmonised data field to denote the patient’s chest pain type during the time of the CT scan.

It holds the following values:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S15: Harmonised values of *chest\_pain\_type*.   | Value | Description | | --- | --- | | 0 | no symptoms | | 1 | typical | | 2 | atypical | | 3 | nonanginal | | 4 | dyspnea | | -1 | unknown | |

Regarding the symptoms: chest pain, dypsnea and other symptoms:

* If a patient has both chest pain (typical, atypical or nonanginal) and dyspnea, chest pain will take the higher priority. Hence, *chest\_pain\_type* will be either 1, 2 or 3
* If a patient has both dyspnea and other symptoms (not chest pain related), dyspnea will take the higher priority. Hence, *chest\_pain\_type* will be 4.
* If a patient has other symptoms that are neither chest pain nor dyspnea, like heart palpitations, *chest\_pain\_type* will be -1.

The general approach is to assume that the patients are asymptomatic (*chest\_pain\_type* = 0) unless indicated that they have a specific type of chest pain (*chest\_pain\_type* = 1, 2 or 3), dypsnea (*chest\_pain\_type* = 4), other symptoms like heart palpitations (*chest\_pain\_type* = -1) or all symptom related data fields are missing (*chest\_pain\_typed* = -1).

*chest\_pain\_type* is harmonised as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S16: Harmonised process of *chest\_pain\_type*.   | Cohort ID | Original Response | Harmonisation Response | | --- | --- | --- | | Cohort A | Column *have\_sob* with  *0* as no.  *1* as yes.  Column *chest\_pain\_type* with  *0* as no chest pain.  *1* as typical chest pain.  *2* as atypcial chest pain.  *3* as nonanginal chest pain. | Map the values of *chest\_pain\_type* and *have\_sob* to *chest\_pain\_type* as follows:  If *chest\_pain\_type* has a value of *1*, *chest\_pain\_type* will be *1*.  Else, if *chest\_pain\_type* has a value of *2*, *chest\_pain\_type* will be *2*.  Else, if *chest\_pain\_type* has a value of *3*, *chest\_pain\_type* will be *3*.  Else, if *chest\_pain\_type* has a value of *0* and *have\_sob* has a value of *1*, *chest\_pain\_type* will be 4.  Else, *chest\_pain\_type* will be 0. | | Cohort B | Column *Dyspnea* with  *no* as no.  *yes* as yes.  Column *Chest Pain Character* with  *no chest pain* as no chest pain.  *typical* as typical chest pain.  *atypcial* as atypcial chest pain.  *nonanginal* as nonanginal chest pain. | Map the values of *Chest Pain Character* and *Dyspnea* to *chest\_pain\_type* as follows:  If *Chest Pain Character* has a value of *typical*, *chest\_pain\_type* will be 1.  Else, if *Chest Pain Character* has a value of *atypical*, *chest\_pain\_type* will be 2.  Else, if *Chest Pain Character* has a value of *nonanginal*, *chest\_pain\_type* will be 3.  Else, if *Chest Pain Character* has a value of *no chest pain* and *Dyspnea* has a value of *yes*, *chest\_pain\_type* will be 4.  Else, *chest\_pain\_type* will be 0. | |

After harmonisation, we validate the values of *chest\_pain\_type* and *symptoms* to ensure that there can only be the following cases:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S17: Valid values of *symptoms* and *chest\_pain\_type*.   | Description | *symptoms* | *chest\_pain\_type* | | --- | --- | --- | | Asymptomatic | 0 | 0 | | Have chest pain | 1 | 1, 2 or 3 | | Only dypsnea | 2 | 4 | | Other symptoms | 2 | -1 | | Unknown | -1 | -1 | |

# References