Case 1.

Use national health insurance claims dataset to answer following questions:
In 2019, compare the difference of CHADS2 score between old people (65 y/o and above) with and without diabetes. (definition: diagnosis code appears in at least 2 outpatient diagnoses, or the diagnosis code appears in any inpatient diagnosis) (only use the first three codes of the ICDs in Table 2)

$CHADS_2$ score or CHA_2DS_2-VASc score are used to predict the risk of ischemic stroke in patients with atrial fibrillation (table 1):

Table 1. $CHADS_2$ score and CHA_2DS_2-VASc score.

	Points assigned	
Risk factors	**$CHADS_2$**	**CHA_2DS_2-VASc**
Age (years)		
65–74		+1
≥75		+2
>75	+1	
Congestive heart failure	+1	+1
Hypertension	+1	+1
Diabetes mellitus	+1	+1
Stroke/TIA	+2	+2
Vascular disease*		+1
Female gender		+1
	Cumulative score: 0–6	Cumulative score: 0–9

*MI, peripheral artery disease or aortic plaque

Lip GY *et al*, 2010.

Table 2. Diagnosis codes used in this study.

Diagnosis	ICD-9-CM codes
Congestive heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4–425.9, 428.x
Hypertension	401.x, 402.x–405.x
Diabetes	250.0–250.3, 250.4–250.9
Ischemic stroke	433.x1, 434.x1, 435.9, 436, 437.1x, 437.9x

Abbreviations: ICD-9-CM, International Classification of Diseases Ninth Revision Clinical Modification system.

Answer:

- Filter people 65 y/o and above, group by Non-Diabetes/diabetes;
- Link outpatient and inpatient datasets
- Define with/without CHF, hypertension, stroke;
- Chi-square test and Wilcoxon Rank-Sum test

Table 1. Distribution of background characteristics and test difference between people with and without diabetes among 65 y/o and above.

	Non-Diabetes (N =6325)		Diabetes (N =2860)		
	N	%	N	%	P
Congestive heart failure	1473	23.29	1138	39.79	<0.0001
Hypertension	2547	40.27	1978	69.16	<0.0001
Ischemic stroke	985	15.57	890	31.12	<0.0001
	Mean	SD	Mean	SD	P
Age	73.89	6.83	73.91	6.43	<0.0001 by ttest; 0.1276by Wilcoxon Rank-Sum test
$CHADS_2$	1.32	1.41	3.09	1.46	<0.0001 by ttest; <0.0001 by Wilcoxon Rank-Sum test

Abbreviations: SD, standard deviation.

If the continuous variables do not follow the normal distribution, non-parametric methods would be used. (in this case Wilcoxon Rank-Sum test)

Case 2.

Use 2016 -2019 inpatient expense datasets (h_nhi_ipdte*), death registry (h_ost_death*), and medical facility dataset (H_nhi_medfa*), for inpatient patients with main diagnosis of ischemic heart disease (ICD9-CM: 410-414), adjust age, gender, diabetes diagnosis (any diagnose code in 250), hospital type, what is the odds ratio (OR) of death of hypertension patients (any diagnose code in 401, 402, 403, 404, 405) compare to non-hypertension.

(*means each year)

(Exclude Hospital type was not medical centers (HOSP_CONT_TYPE: 1, 2, 3))

Answer:

Analysis Process:

- Filter ischemic heart disease (ICD9-CM: 410-414) in each year
- Filter HOSP_CONT_TYPE: 1, 2, 3 (3 as reference) only include patients in hospital type 1,2,3
- Exclude: hospitalized before January 1, 2016 or have not been discharged after December 31, 2019; exclude unknow gender
- Link (h_ost_death*), only include people that have death date, and the admission data <= death date<= 3 days after discharge, will be considered as death due to ischemic heart disease.

OR= 2.315, means among ischemic heart disease inpatient patients, the odds of death among patients with hypertension is 2.315 times the odds of death among non-hypertension patients. However, the 95% confidence interval is 0.205 to 26.173, considering no statistical significance.