

Quality Assessment and Validation of Administrative Data Sources in Health Statistics

21 - Administrative Data Systems

2.6.2016

Tina Zupanič & Metka Zaletel

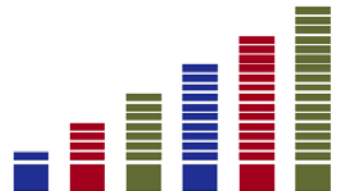
National Institute of Public Health,

Ljubljana, Slovenia

tina.zupanic@nijz.si; metka.zaletel@nijz.si

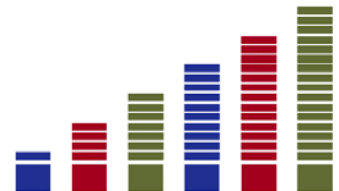
Content

- Administrative data sources in health statistics
- Integration Phase
- Quality Drawbacks of administrative data sources
- Results and case studies
- Conclusions

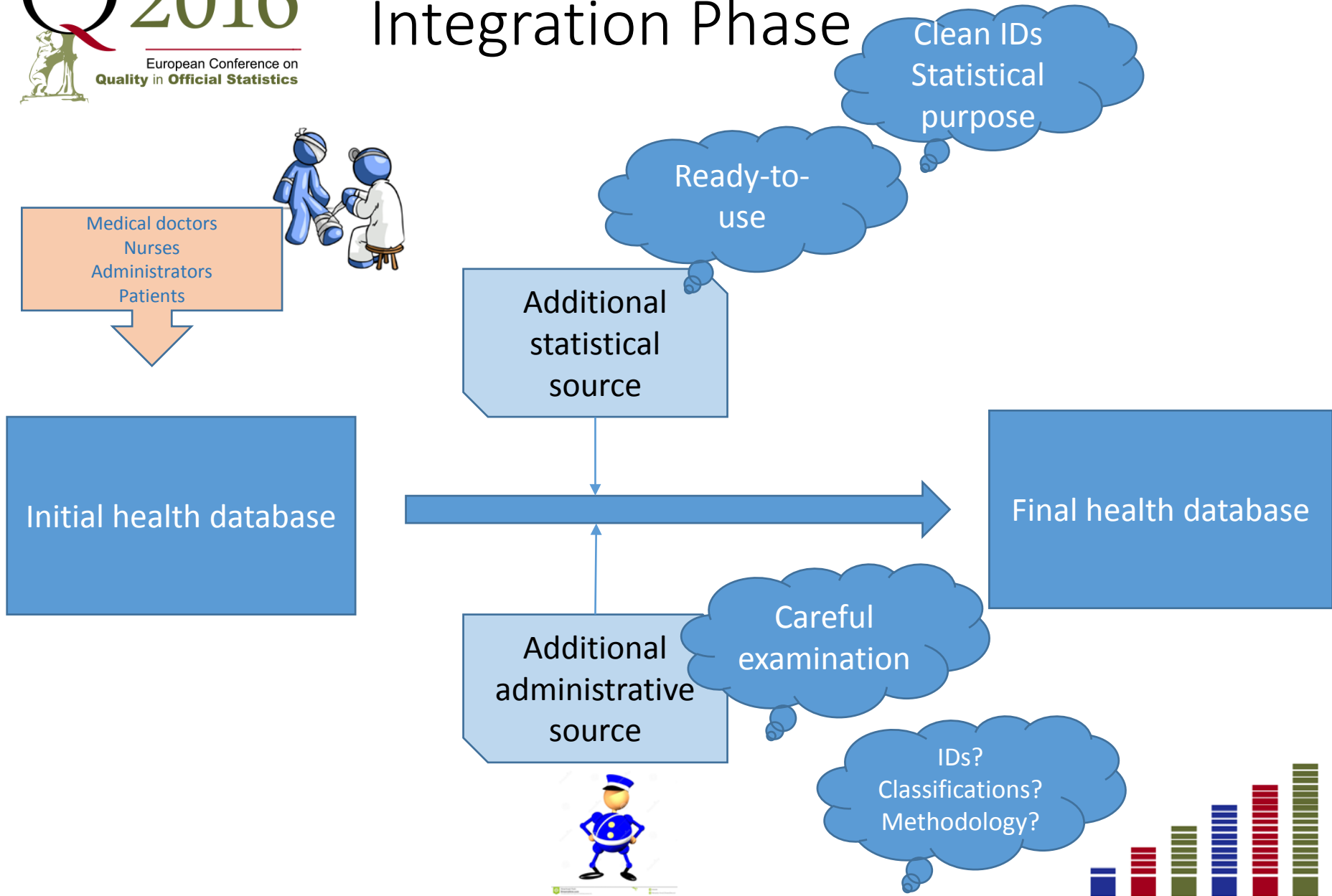


Administrative data sources in health statistics

- Administrative data sources (ADS) are traditionally used in the process of preparation of health statistics.
- Almost no additional ADS have been used in the process with the role of complementary data source or the role of improving the quality of the final estimations.
- Health statistics are usually produced by institutions other than NSIs that are sometimes not very obviously a part of the system of national statistics.
- Hard to apply the legal framework of official statistics and gather administrative data sources to raise the quality level of the original data.



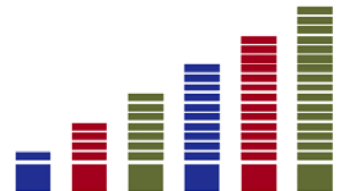
Integration Phase



Quality Drawbacks of ADS

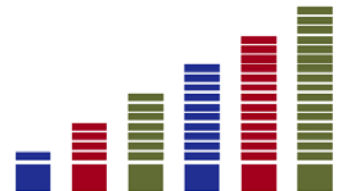
- Quality drawbacks exist in initial health database and in additional ADS.
- Initially caused by:
 - Lack of knowledge or interest in final data (e.g. administrators at the hospitals)
 - Lack of time (e.g. medical doctors and nurses)
 - Stress during the process of data collection (e.g. patients)
- In the additional ADS, the drawbacks are caused by:
 - Different definitions
 - Different classifications
 - Different time span

in comparison with the initial database.



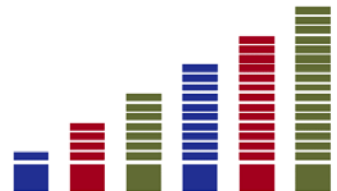
Results and case studies

1. First case study - poor quality of socio-demographic data
2. Second case study - causes of death (CoD) data
3. Third case study – police transport accidents



First case study - poor quality of socio-demographic data

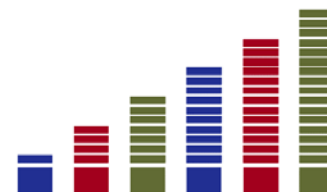
- in the National Hospital Health Care Statistics Database (hospitalization database) and the Perinatal Information System (perinatal database)
- linking with population and census socio-demographic data gathered by the SORS
- on the basis of a personal identification number



Marital status

Tab. 1: Marital status in the perinatal database, original data vs. official data (NIJZ, 2014)

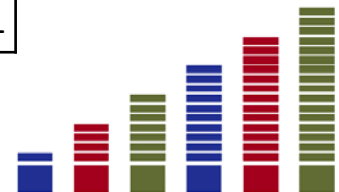
de facto marital status	2014	legal marital status	2014
single	11.4	single	55.6
married	19.7	married	42.1
divorced	0.2	widowed	0.1
widowed	0.0	divorced	2.1
living in consensual unions	29.0	same-sex registered partnership	0.0
unknown/missing	39.7	unknown/missing	0.1
Total	100	Total	100



Highest level of education completed

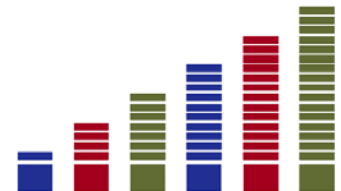
Tab. 2: The highest level of education completed, a comparison of the original (HDD and PIS) vs. official data (NIJZ, SORS 2013 and 2014)

	HDD vs. official data				PIS vs. official data			
	2013		2014		2013		2014	
	SBO	SORS	SBO	SORS	PIS	SORS	PIS	SORS
valid value	23.5	80.9	28.5	80.8	67.1	99.0	49.7	98.9
unknown/missing	76.5	19.1	71.5	19.2	32.9	1.0	50.3	1.1



Second case study - causes of death (CoD) data

- changed methodology in capturing CoD in hospitals from 2013 onward
- underlying CoD for those individuals that died during hospitalization was linked from annual CoD database, based on personal identification numbers
- from 2013 onward hospital data capture direct CoD and not an underlying CoD

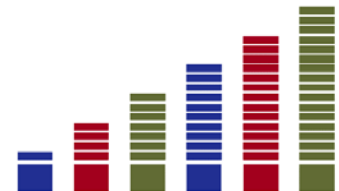


Underlying/ direct CoD for hospitalized individuals – captured vs. official data

Chapter of Direct/ Underlying CoD	Year of data capture						
	Underlying/ Direct CoD					Underlying CoD	
	2010	2011	2012	2013*	2014*	2013	2014
1 Certain infectious and parasitic diseases	7.53	9.37	9.08	8.57	6.96	0.91	1.31
2 Neoplasms	21.65	21.2	18.42	17.52	19.57	37.8	36.85
3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	0.2	0.31	0.28	0.23	0.2	0.17	0.15
4 Endocrine, nutritional and metabolic diseases	0.42	0.54	0.63	0.44	0.62	1.99	2.04
5 Mental and behavioural disorders	0.11	0.08	0.09	0.19	0.11	0.79	0.74
6 Diseases of the nervous system	1.03	0.9	1.08	0.86	0.97	1.45	1.62
7 Diseases of the eye and adnexa	0	0.01	0	0	0	0	0
8 Diseases of the ear and mastoid process	0	0	0	0	0	0.01	0
9 Diseases of the circulatory system	35.25	34.85	35.66	37.86	38.61	33.65	36.15
10 Diseases of the respiratory system	18.64	19.21	20.9	19.37	18.93	9.04	8.29
11 Diseases of the digestive system	4.25	5.07	4.84	3.24	3.12	8.8	8.1
12 Diseases of the skin and subcutaneous tissue	0.04	0.11	0.12	0.12	0.05	0.25	0.08
13 Diseases of the musculoskeletal system and connective tissue	0.11	0.16	0.15	0.11	0.03	0.49	0.63
14 Diseases of the genitourinary system	1.09	1.61	1.46	1.19	1.2	2.32	1.94
15 Pregnancy, childbirth and the puerperium	0	0	0	0	0.01	0	0.01
16 Certain conditions originating in the perinatal period	0.01	0	0.02	0.03	0.01	0.01	0.01
17 Congenital malformations, deformations and chromosomal abnormalities	0	0.01	0.01	0.03	0.03	0.21	0.26
18 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	9.54	6.42	7.06	9.85	9.27	0	0.01
19 Injury, poisoning and certain other consequences of external causes	0.12	0.14	0.13	0.15	0.18	0	0
20 External causes of morbidity and mortality	0.01	0	0	0	0	2.09	1.82
21 Factors influencing health status and contact with health services	0	0	0.04	0.23	0.09	0	0

Third case study – police transport accidents

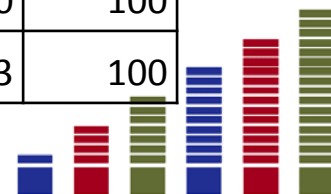
- police data were used for the identification of all cases of traffic accidents
- police transport database was linked to in-patient data
- in-patient data include only more severe injuries (those that needed to be hospitalized) and not those which were only treated in EDs'
- before linking - transformation of ICD-10 codes of injuries and poisonings to Abbreviated Injury Scale (AIS) codes
- no personal identification number available in police transport database (linkage according to the three common variables: date of birth, date of accident/date of injury and gender)
- hospitalization data were assessed as an *inadequate source* for the severity of injury



Severe transport injury – linkage outcome

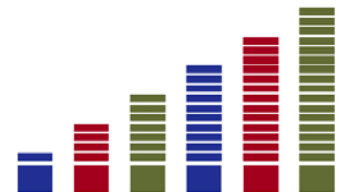
Tab. 4: The outcome of linking police transport database and in-patient database, row percentages (Police and NIJZ, 2013)

		AIS - 3 categories			Total
		AIS 1, 2 (less severe injury)	AIS 3, 4, 5, 6 (severe injury)	cannot determine	
Accident outcome (police data)	missing	100.0	0.0	0.0	100
	material damage	86.2	6.9	6.9	100
	with severe body injury	76.2	22.3	1.4	100
	with less severe body injury	98.0	1.4	0.6	100
	fatal accident	33.3	57.6	9.1	100
	without injury	88.2	11.8	0.0	100
Total		90.0	8.7	1.3	100



Conclusions

- Administrative (or statistical) data sources with approved quality level → caution needed due to different methodology, but still used to improve the final estimates
- Administrative data sources (non-statistically edited earlier in the process) → in some cases useless due to different purpose of collection or some other reason
- Health statistics **can gain** from additional statistical and administrative data sources, but the process of validation of different additional sources has begun recently.
- Lots **can be learnt** from business statistics.
- **Big data** creates another possibility to enrich the existing databases, but again – in the case of health statistics, it has not been explored enough.



Thank you!

