Vor dem Gesetz sind alle gleich – und bei der Datenerhebung?

19. Deutsch-Österreichisch-Schweizerisches Demografentreffen
Wien

24. Oktober 2019

Matthias Bopp

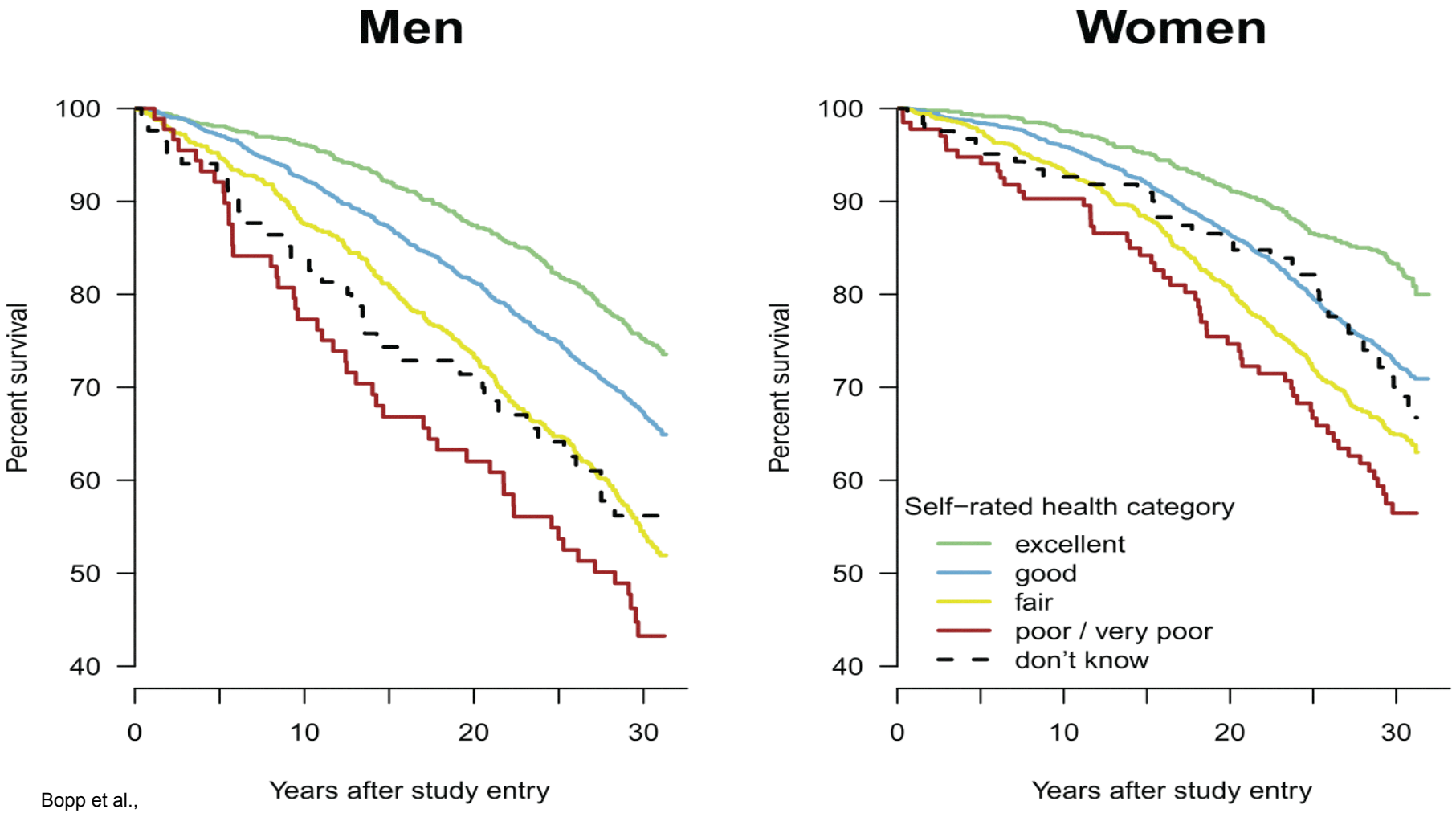
Vorgeschichte

Von der Datenerhebung zur Auswertung:
Individuum – Registration – Aufbereitung – Analyse

Keine Angabe / Missing values

= keine Information ?

Mortalität nach selbst berichteter Gesundheit



Bopp et al.,
PLoS ONE 2012

Unbegründete Variation nach

- Statistikjahr
- Geschlecht
- Alter
- Region

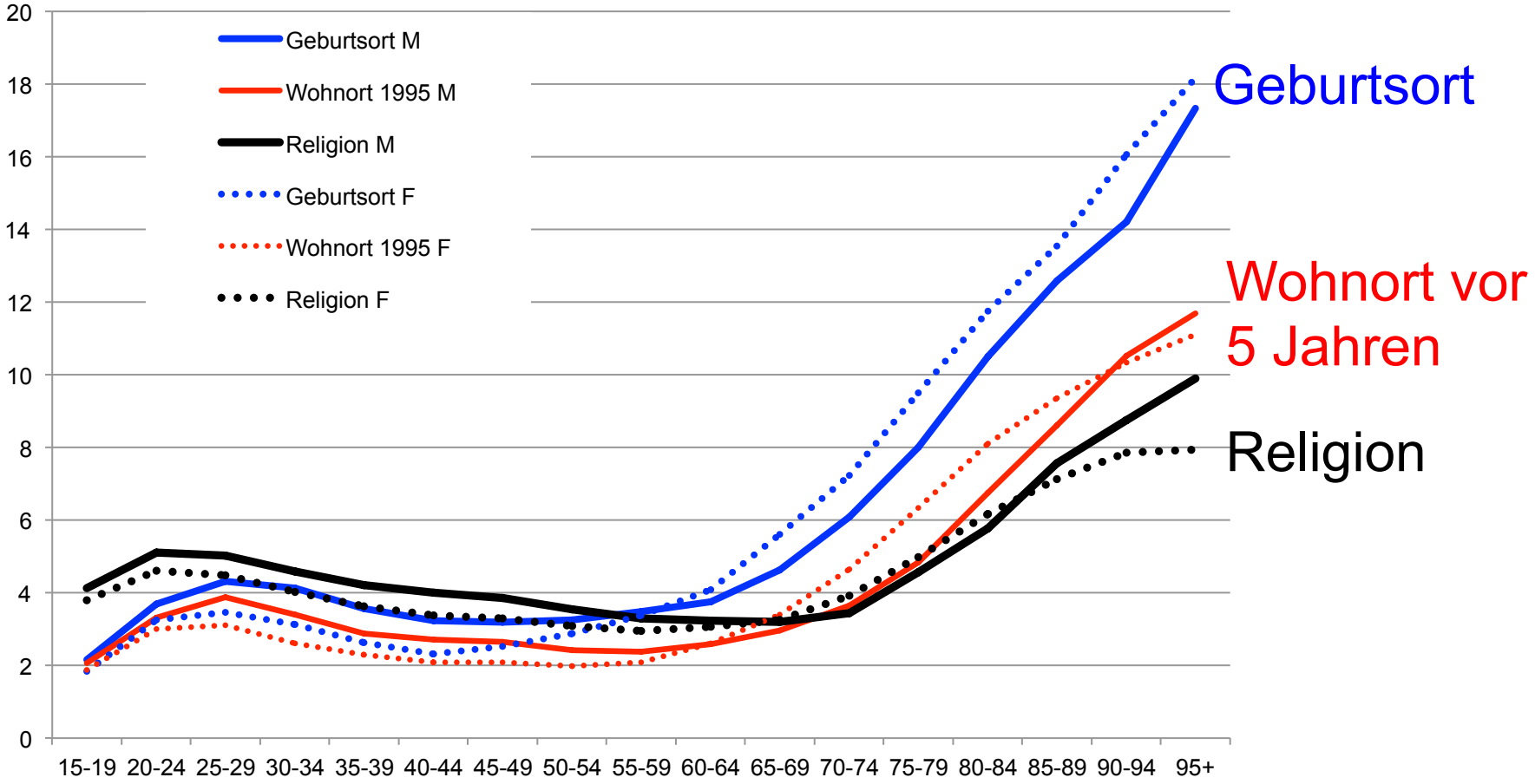
Beispiel: Schweizerische Volkszählung 2000,
(N=6'037'986 15+-Jährige / 5‘616‘616 20+Jährige)

Keine Angabe für:

- Geburtsort (Gemeinde, Staat): 4.15%
- Wohnort vor 5 Jahren: 3.1%
- Religion: 4.0%
- Freiwillige Tätigkeit: 24.4%
- Schulabschluss: 8.3%
- Erlernter Beruf: 37.7%

Fehlende Angaben nach Alter und Gechlecht

Volkszählung 2000, N=6'037'986



Fehlende Angabe zum Geburtsort: Logist. Regression

N=5616616 ab 20-Jährige		Odds Ratio	[95% CI]
Altersklasse	20-24	(ref.)	
	25-29	1.33	1.30-1.36
	30-34	1.25	1.26-1.32
	35-39	1.05	1.02-1.07
	40-44	**0.92**	**0.90-0.94**
	45-49	**0.95**	**0.93-0.97**
	50-54	1.05	1.03-1.08
	55-59	1.18	1.15-1.21
	60-64	1.11	1.09-1.14
	65-69	1.21	1.18-1.24
	70-74	1.39	1.36-1.42
	75-79	1.69	1.65-1.73
	80-84	2.15	2.10-2.21
	85+	**2.90**	**2.83-2.98**
Geschlecht	Mann	(ref.)	
	Frau	**0.82**	**0.81-0.83**
Bildungsstand	**K.Angabe**	**13.96**	**13.80-14.13**
	tief	(ref.)	
	mittel	0.53	0.52-0.53
	tertiär	0.38	0.37-0.39
	Universität	**0.23**	**0.23-0.24**

Fehlende Angabe zur Freiwilligen-Tätigkeit: Logist. Regression

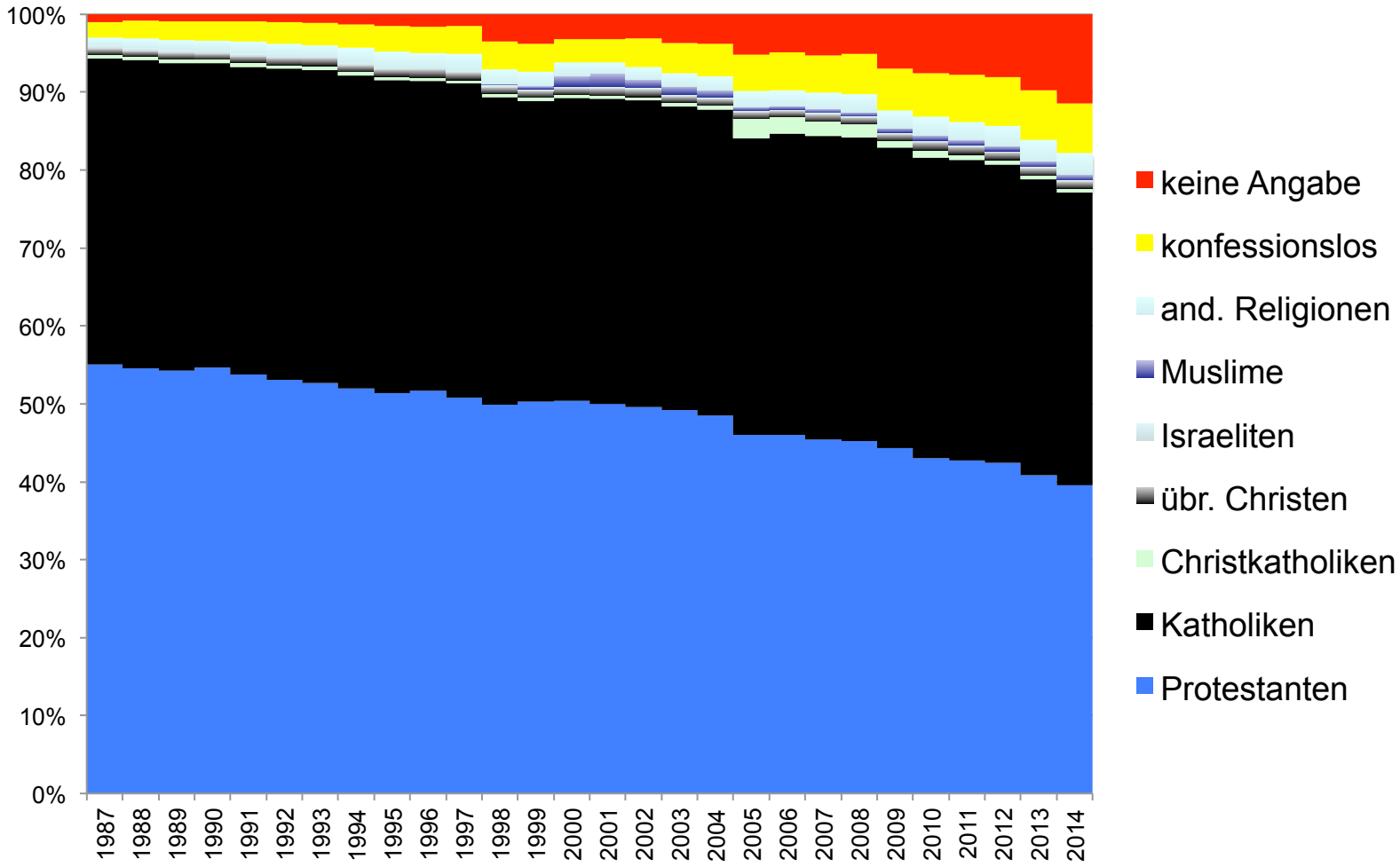
N=5616616 ab 20-Jährige		Odds Ratio	[95% CI]
Altersklasse	20-24	(ref.)	
…			
	35-39	0.77	0.76-0.78
	40-44	0.76	0.75-0.76
…			
	85+	2.15	2.12-2.18
Geschlecht	Mann	(ref.)	
	Frau	0.59	0.59-0.60
K. Angabe:	**Bildungsabschluss**	**5.19**	5.16-5.24
K. Angabe:	**Religion**	**3.39**	3.35-3.44
K. Angabe:	**Geburtsort**	**2.95**	2.92-2.99
K. Angabe:	**Wohnort 1995**	**2.68**	2.65-2.73
Grossregion	Léman	(ref.)	
	Espace Mittelland	0.72	0.71-0.72
	Nordwest-CH	0.67	0.66-0.67
	Zürich	0.68	0.67-0.68
	Ost-CH	0.68	0.67-0.68
	Zentral-CH	0.69	0.68-0.69
	Ticino	1.16	1.15-1.17

Restgruppen

= vernachlässigbare Information ?

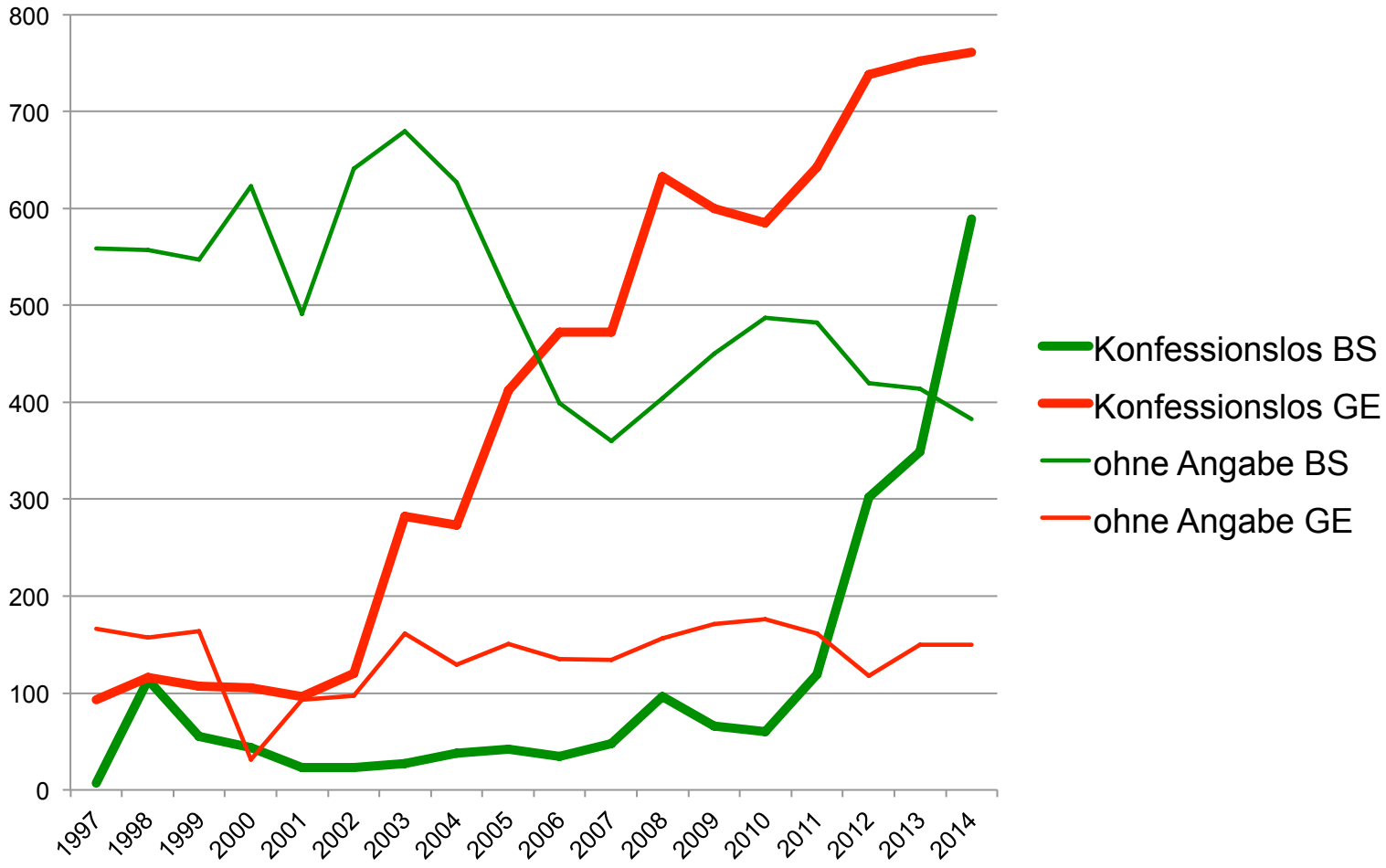
Religionszugehörigkeit: 1987-2014 in der Schweiz Verstorbene

Statistik der Todesfälle und -ursachen



Religionszugehörigkeit: regionale Unterschiede

Statistik der Todesfälle und -ursachen, 1997-2014



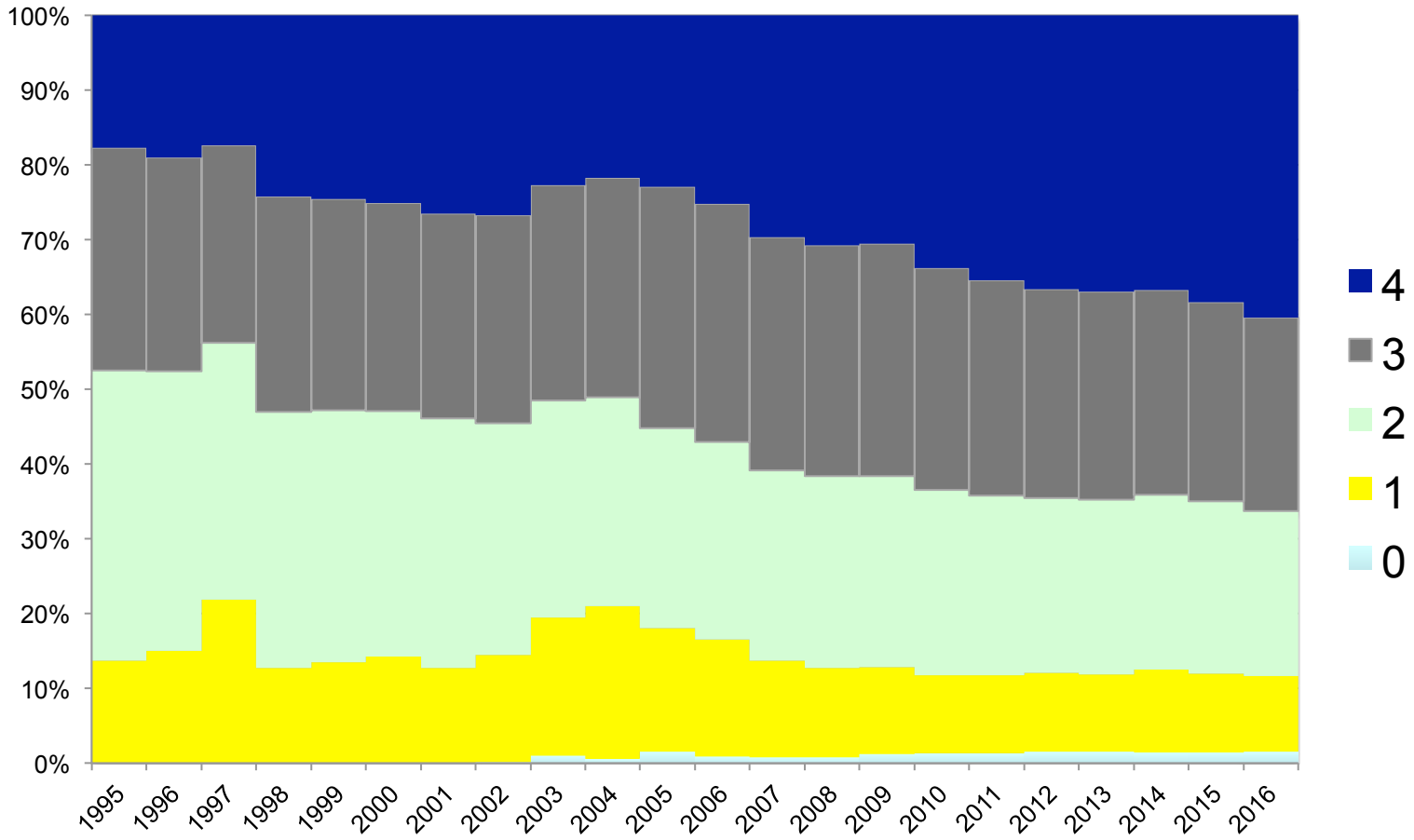
Todesursachenstatistik in der Schweiz

1) Arzt füllt max. 4 Felder für Todesursache aus:
   - zugrundeliegende
   - unmittelbare
   - max. 2 Begleiterkrankungen

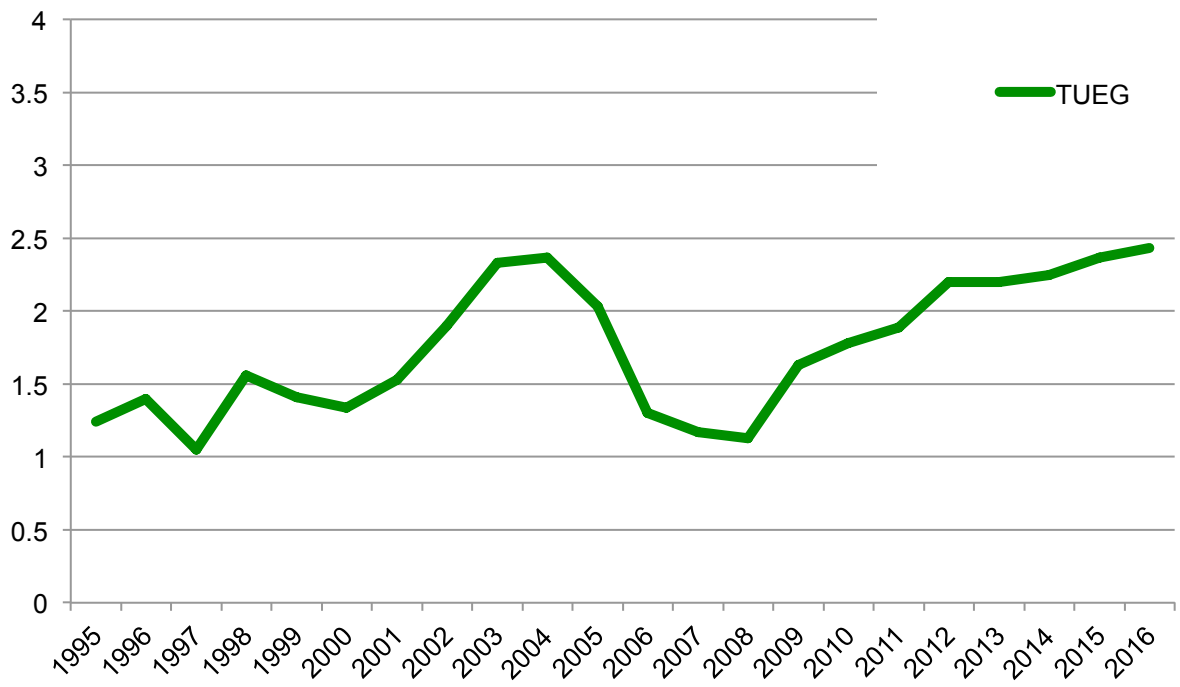
2) BFS bestimmt die „endgültige“ Todesursache
   (massgeblich für alle offiziellen Statistiken)

Anzahl Todesursachen

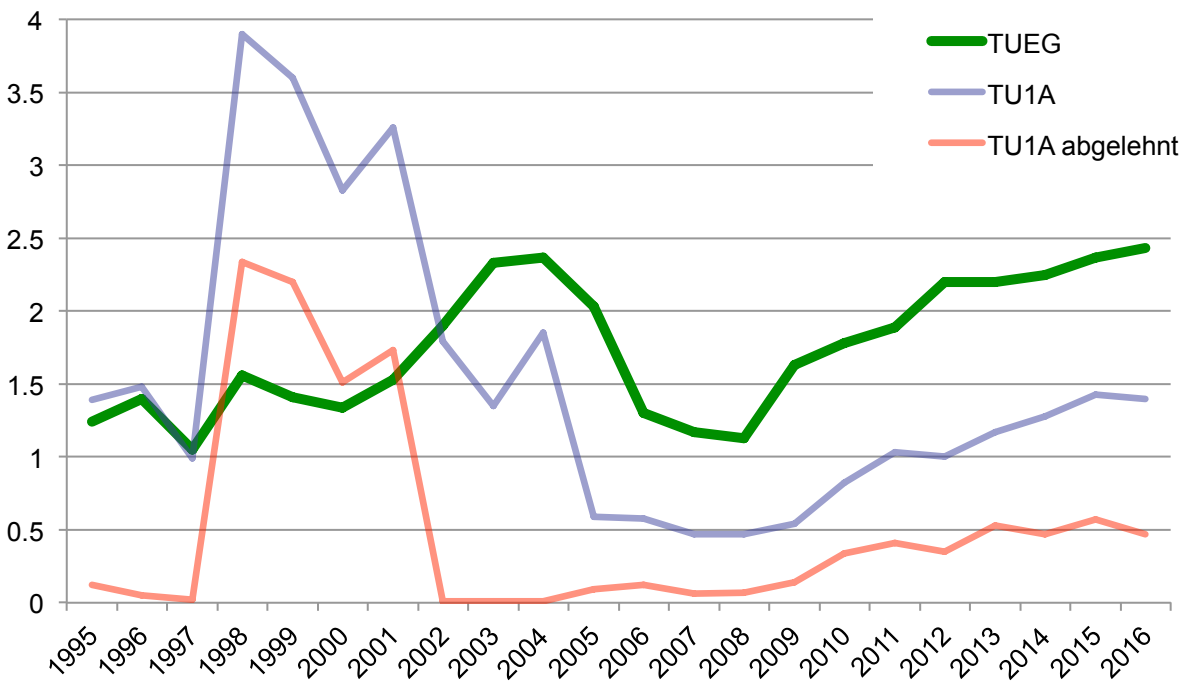
Statistik der Todesfälle und -ursachen 1995-2016



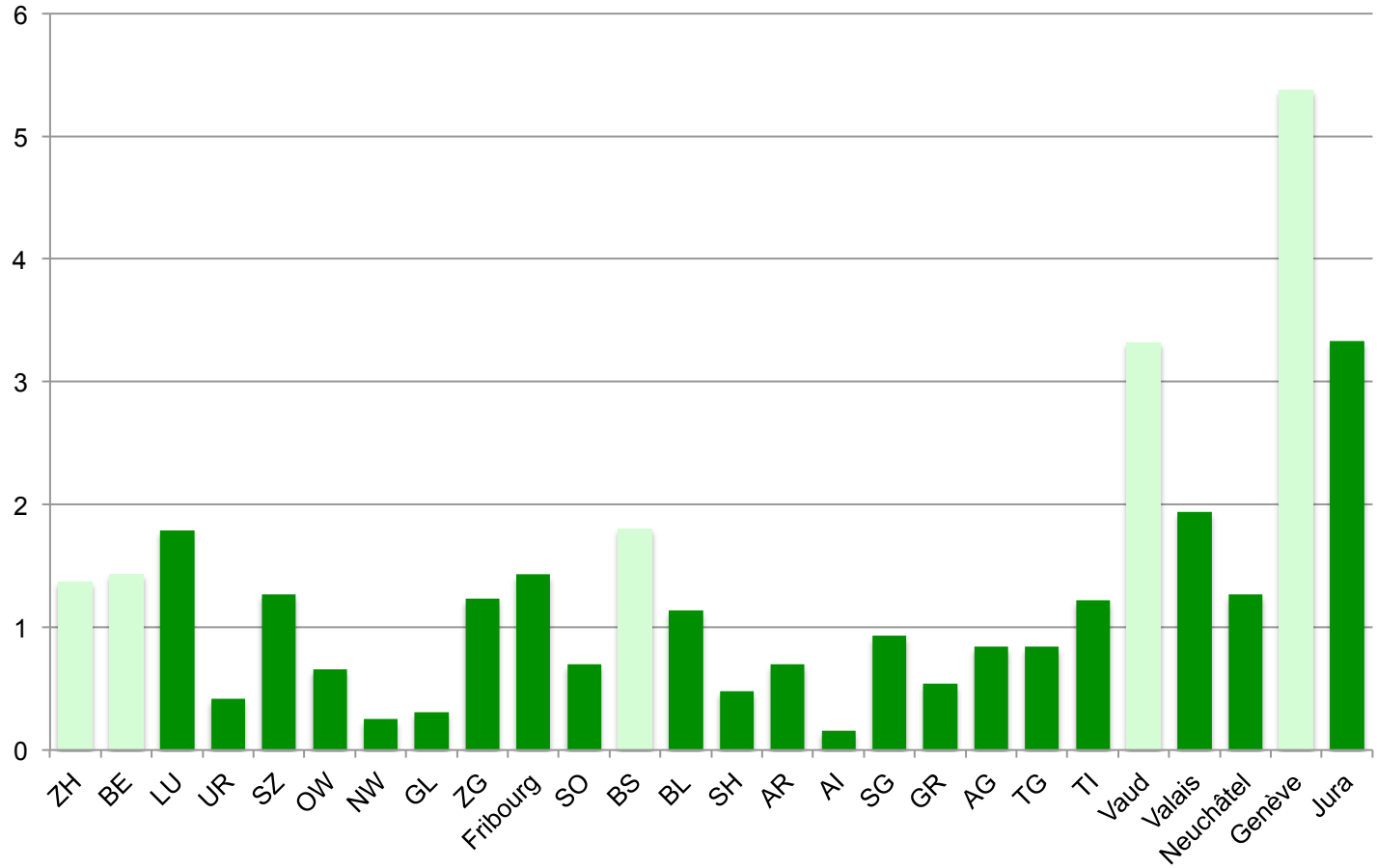
Anteil (%) **Todesursache unbekannt**: TUEG nach Statistikjahr



Anteil (%) **Todesursache unbekannt**: TUEG (BFS) vs. TU1A (Original) nach Statistikjahr

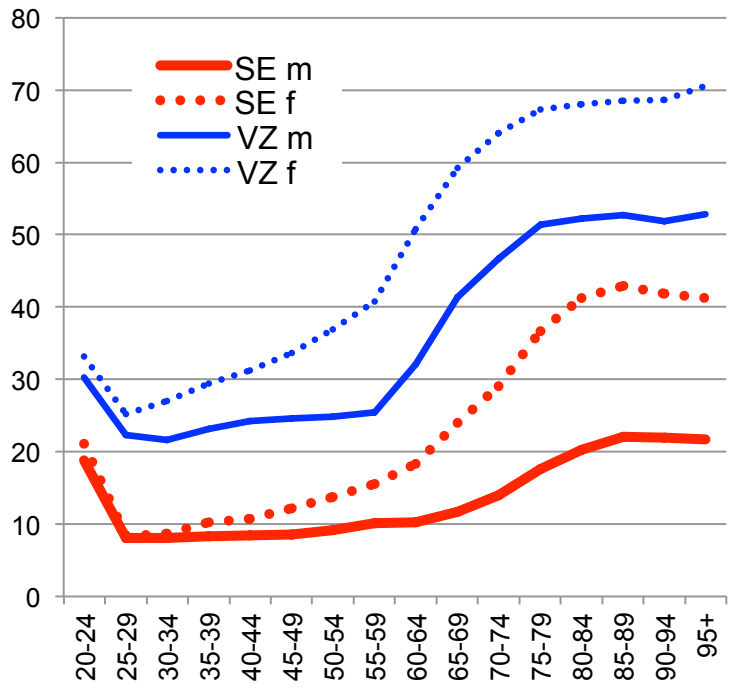


Anteil (%) **Todesursache unbekannt** (TUEG): 65+-Jährige nach Kanton

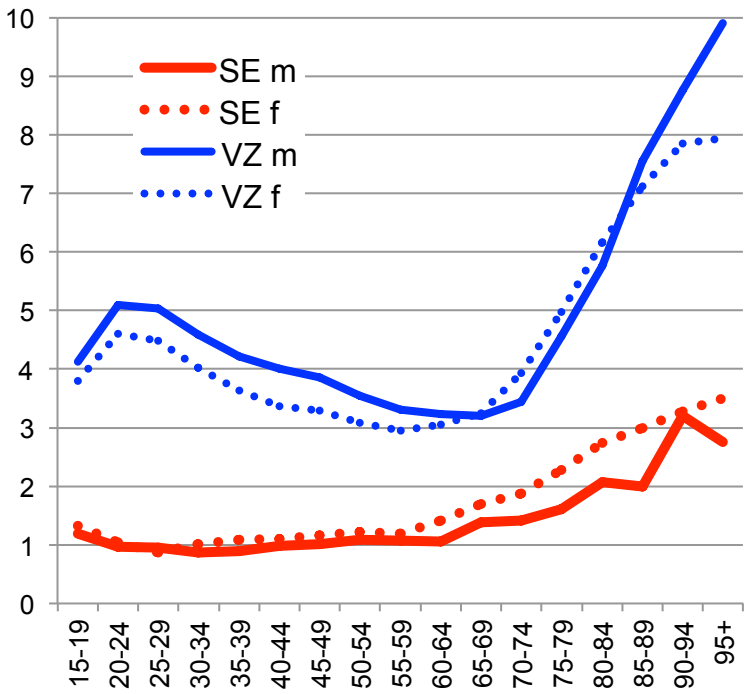


Strukturerhebung 2013-17 vs. Volkszählung 2000: Fehlende Angaben für Erlernten Beruf bzw. Religion

Erlernter Beruf



Religionszugehörigkeit



Schlussfolgerungen

- Fehlende Angaben sind keine informationsleeren Zufallsereignisse, sondern häufig ein Indiz für schlechtere Datenqualität

- Sprünge in den Daten sind
  (fast) immer verdächtig !

Letzte Dezimale Körpergrösse SCHWEIZ

BMC Public Health 2008, **8**:342 http://www.biomedcentral.com/1471-2458/8/342

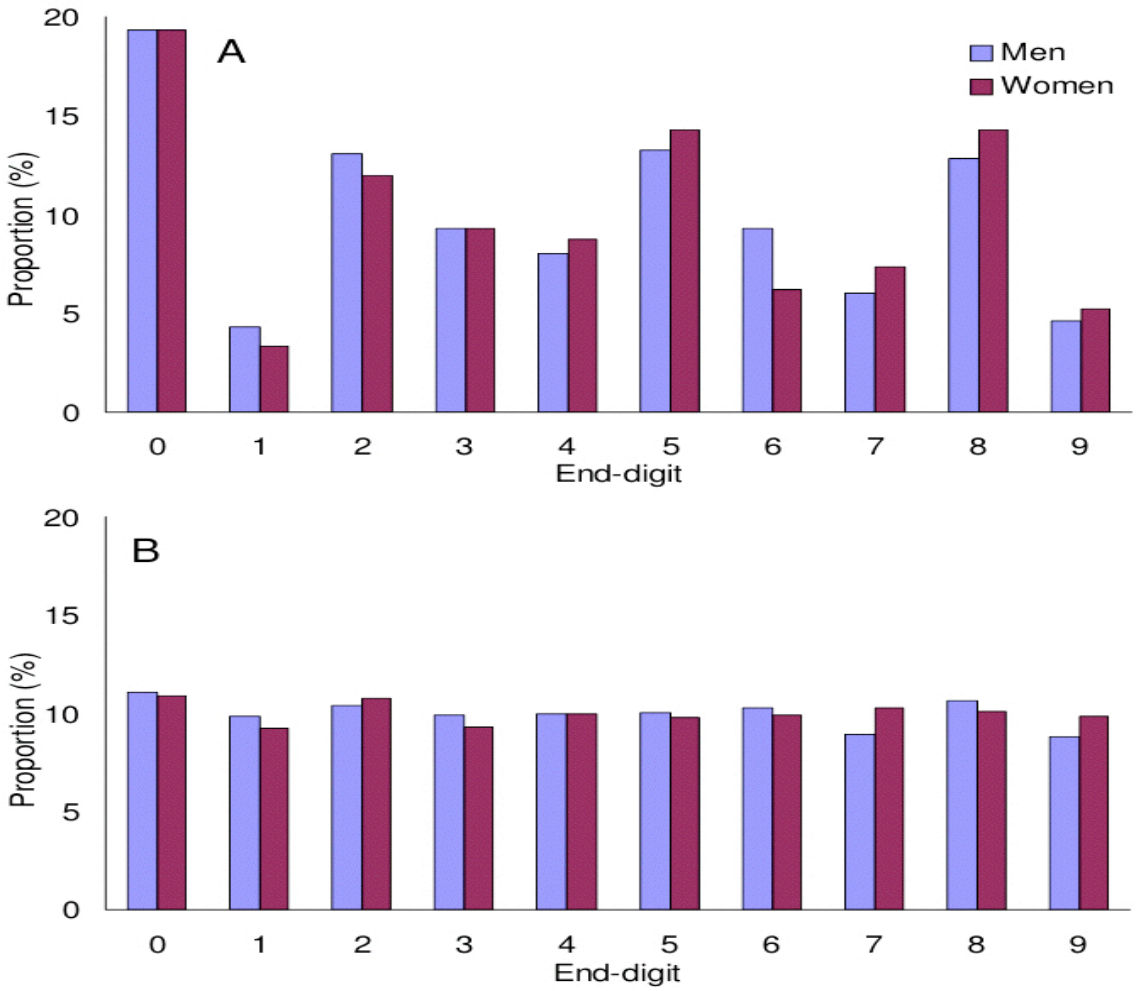
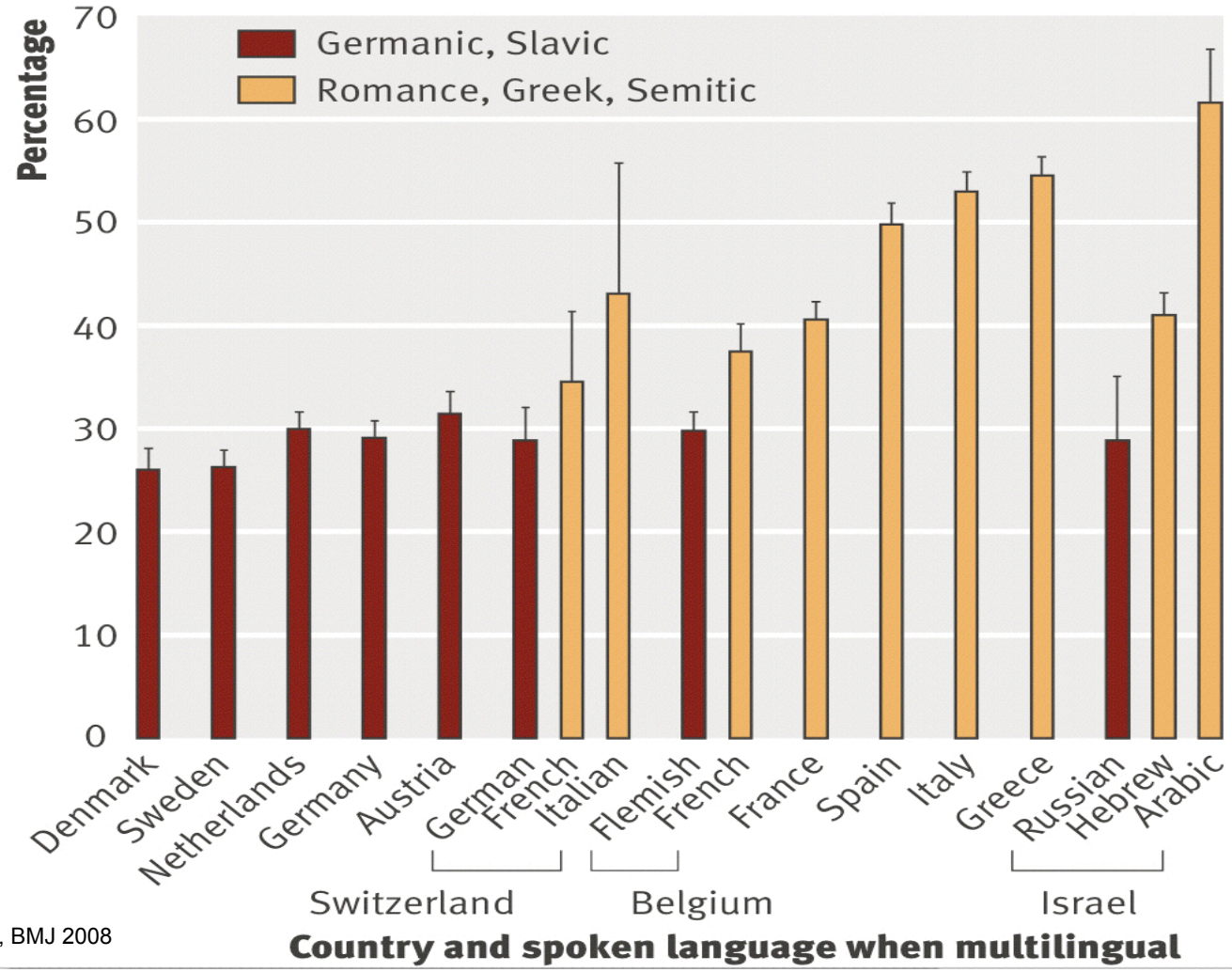


Figure 2
Distribution of end-digits of height (cm) by sex. A: Swiss Health Surveys 1992–2002, (self-reported, N = 47,192), B: MONICA study, Switzerland 1984–1993, (measured, N = 10,144). MONICA: Monitoring Trends and Determinants in Cardiovascular Disease.

Bopp & Faeh, BMC Public Health 2008

Letzte Dezimale Körpergrösse 0 oder 5: international

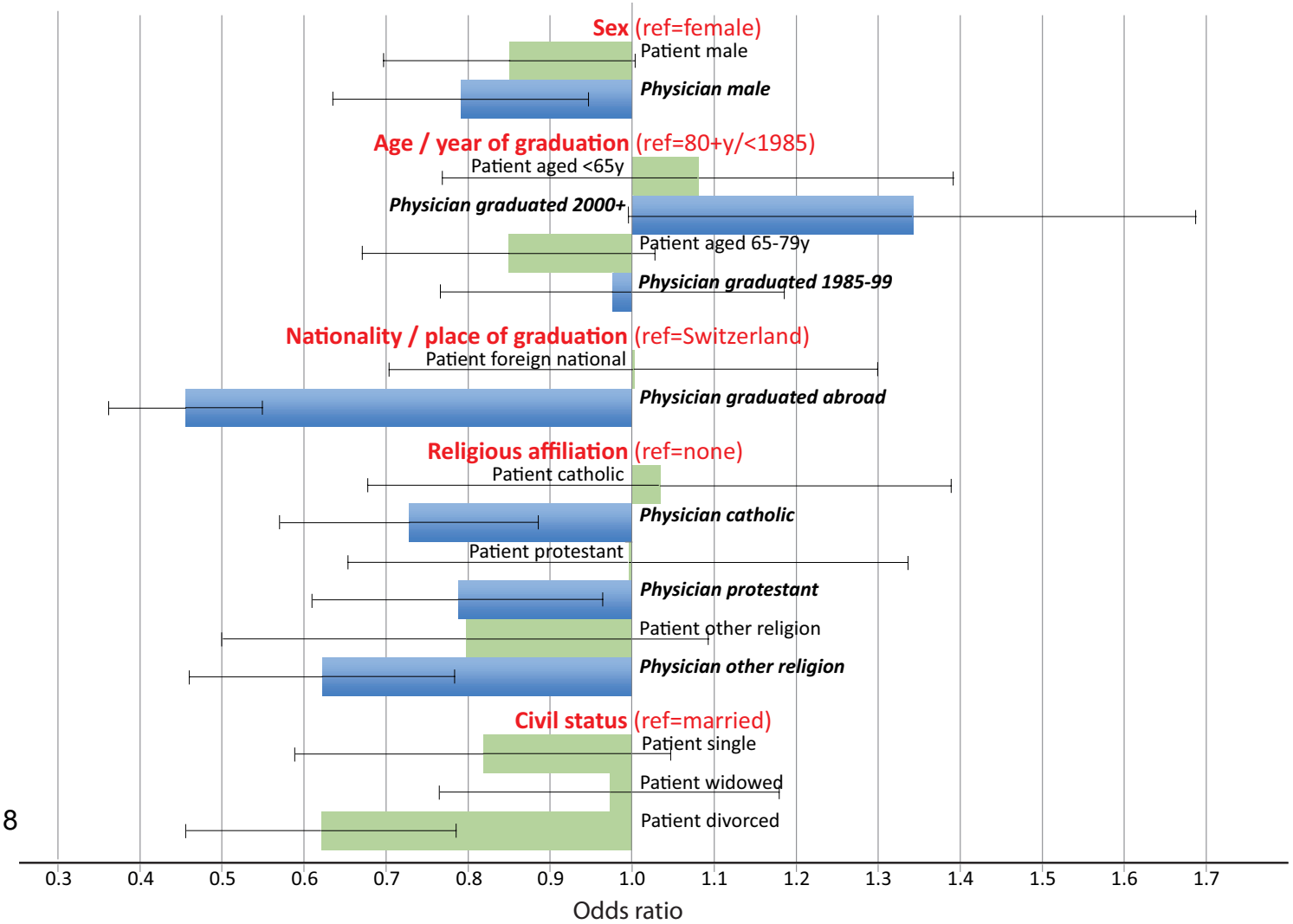


Proportion of study participants using the end digits zero or five when reporting body height (with upper 95% CI)

Bopp & Faeh, BMJ 2008

- Einstellungen und Prägungen können ein langes Nachleben haben

Medizinische Entscheidungen am Lebensende: Patienten- vs. Arztmerkmale, Schweiz 2013/14



Bopp et al.,
PLoS ONE 2018

- es gibt keine Patentrezepte zum Umgang mit fehlenden Angaben oder unerklärten Diskontinuitäten in Datenreihen !