

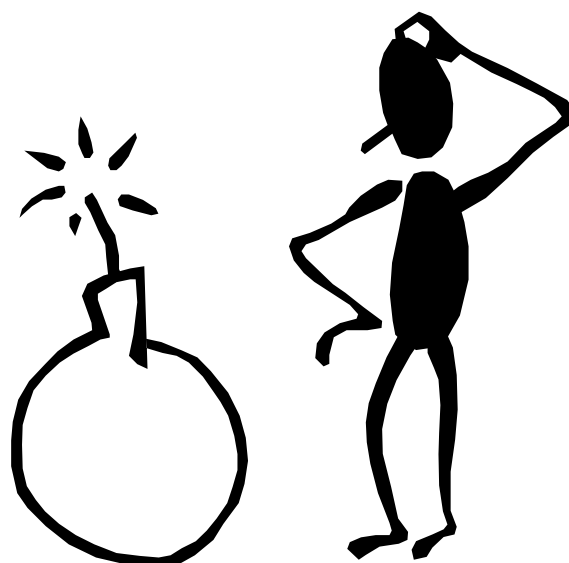


# Lecture 10

## Presentation of results



## Data available – how to proceed?





## Which ways do you know to present data?



## Summarizing categorical data: tables

- Presentation in frequency tables:
- count events in every category
- report absolute frequencies
- report relative frequencies

**Table 3.1** Method of delivery of 600 babies born in a hospital.

Method of delivery	No. of births	Percentage
Normal	478	79.7
Forceps	65	10.8
Caesarean section	57	9.5
Total	600	100.0

Tables have informative legends, column and row titles, informative entries, footnotes





## Summarizing categorical data: bar chart

Length of the bar represents frequency

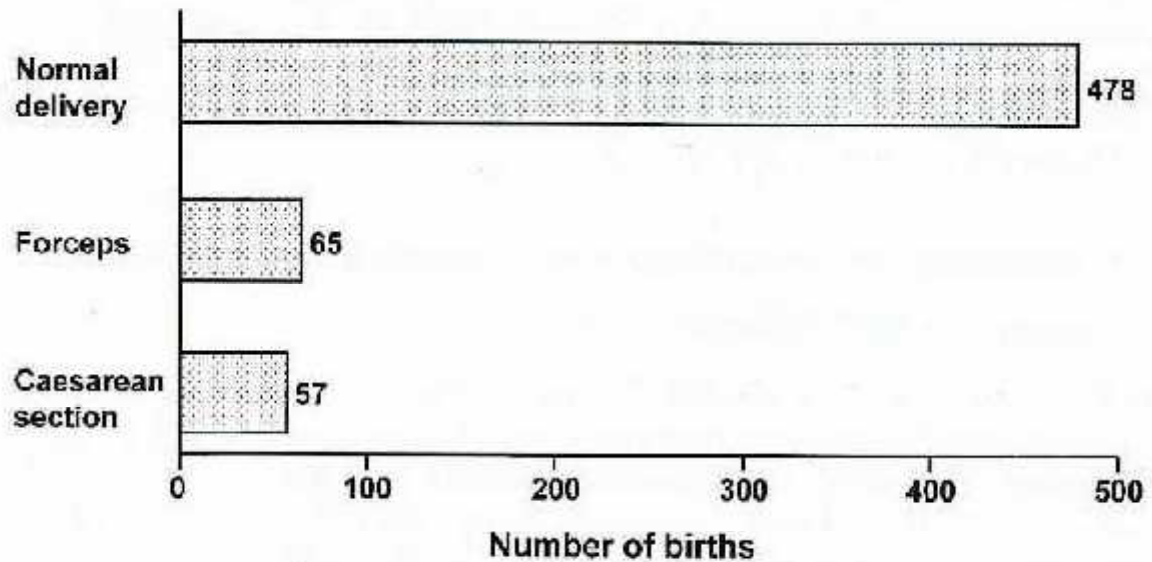


Fig. 3.1 Bar chart showing method of delivery of 600 babies born in a hospital.



## Summarizing categorical data: pie chart

Area of the sector represents frequency  
relative frequency (angle =  $360^\circ \times \text{relative frequency}$ )

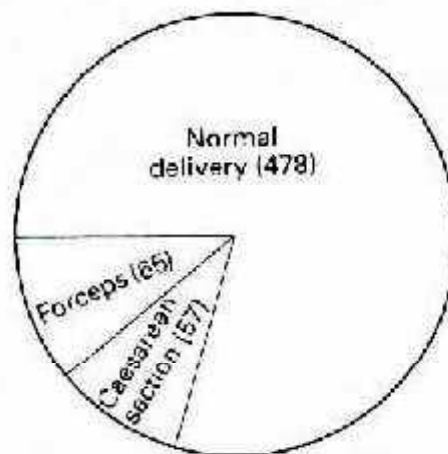


Fig. 3.2 Pie chart showing method of delivery of 600 babies born in a hospital.

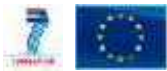


## Summarizing categorical (ordinal) data: tables

- Nearly identical to the nominal presentation.
- Important: ordering!
- It makes sense to additionally report the cumulative sums.

<i>packyears</i>	<i>count</i>	<i>percent</i>	<i>cumulative count</i>	<i>cumulative percent</i>
<i>0</i>	203	26.4	203	26.4
<i>0 &lt;..<lt; 20<="" i=""></lt;></i>	297	38.6	500	65.0
<i>20 &lt;=..<lt; 40<="" i=""></lt;></i>	147	19.1	647	84.1
<i>40 &lt;=..</i>	122	15.9	769	100.0

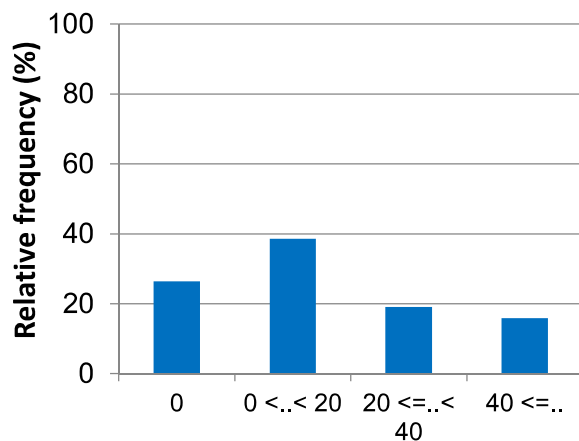
in tables...



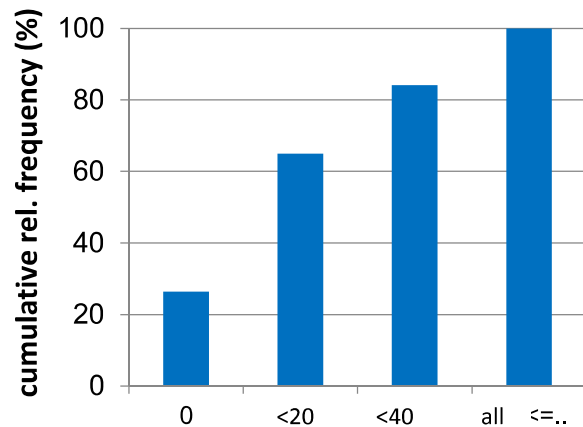
## Summarizing categorical (ordinal) data: bar chart

... and graphs.

bar chart



cumulative bar chart



Smoking categories (packyears)





## The "Acid Test" for Tables and Figures

- Any Table or Figure must be sufficiently clear, well-labeled, and described by its legend to be understood by your intended audience without reading the results section
  - it must be able to stand alone and be interpretable.
- Overly complicated Figures or Tables may be difficult to understand - strive for simplicity whenever possible.
- If you are unsure whether your tables or figures meet these criteria, give them to a colleague and ask them to interpret your results.



## Essential table details

Variable labels

Labels of variable values

Table legend/caption

Column titles

Table body (data)

Table 1. Population totals and proportions of older adults for the world and by INDEPTH country, in 2009 and projected to 2030

Country	Country income category <sup>a</sup>	2009			2030		
		Total, N <sup>b</sup>	50+, N (%)	60+, N (%)	Total, N	50+, N (%)	60+, N (%)
World		6,829	1,379 (20.2)	737 (10.8)	6,309	2,288 (36.3)	1,370 (21.5)
Sub-Saharan Africa		843	110 (13.0)	54 (6.3)	1,305	157 (12.0)	78 (5.9)
South Africa	UMI	63	8 (12.7)	4 (6.3)	55	10 (18.2)	6 (11.1)
Tanzania	Low	44	4 (9.1)	2 (4.5)	75	8 (10.7)	4 (5.3)
Kenya	Low	40	3 (7.5)	2 (5.0)	63	7 (11.1)	3 (4.8)
Ghana	Low	24	2 (8.3)	1 (4.2)	35	5 (14.3)	3 (8.6)
Asia		4,121	765 (18.6)	400 (9.7)	4,917	1,395 (28.4)	821 (16.7)
Viet Nam	Low	88	15 (17.2)	6 (6.8)	105	32 (30.5)	19 (18.2)
Bangladesh	Low	162	29 (17.9)	13 (8.0)	203	46 (22.7)	25 (12.3)
Indonesia	UMI	230	40 (17.4)	20 (8.7)	271	76 (28.0)	45 (16.6)
India	UMI	1,198	187 (15.6)	89 (7.4)	1,485	343 (23.1)	185 (12.4)
Footed (INDEPTH country id) totals		1,305	201 (15.3)	135 (10.3)	2,293	531 (23.2)	286 (12.5)

<sup>a</sup>World Bank country income category: Low, low income; UMI, lower-middle income; UMI, upper-middle income.

<sup>b</sup>N in millions (1,000,000).

Sources: UN Population Division (15) and World Bank (16).



footnotes





## Key points - tables

- Use clear and informative titles (above the table)
- Use a separate cell for each value
- Use short, descriptive row and column titles
- Use only horizontal line borders
- Always include the units, error values and number of samples
- Round off the numbers so that readers can make meaningful comparisons more easily
- Consider summary measures (mean, SD) for continuous data



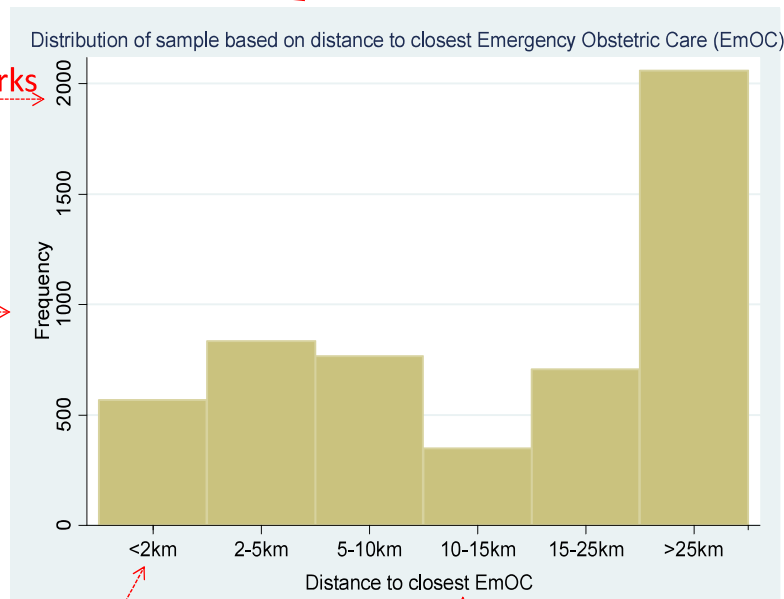
Figure title/legend

Major tick marks

Y-axis label

Category labels

X-axis label







## Key points – figures/graphs

- Use clear and informative titles (below the figure/graph)
- Choose the right figure for the variable
- Pie charts can be difficult to interpret/understand
- Clearly label axes, categories, and major tick marks
- Black and white or colour (?)



## Figures - Example

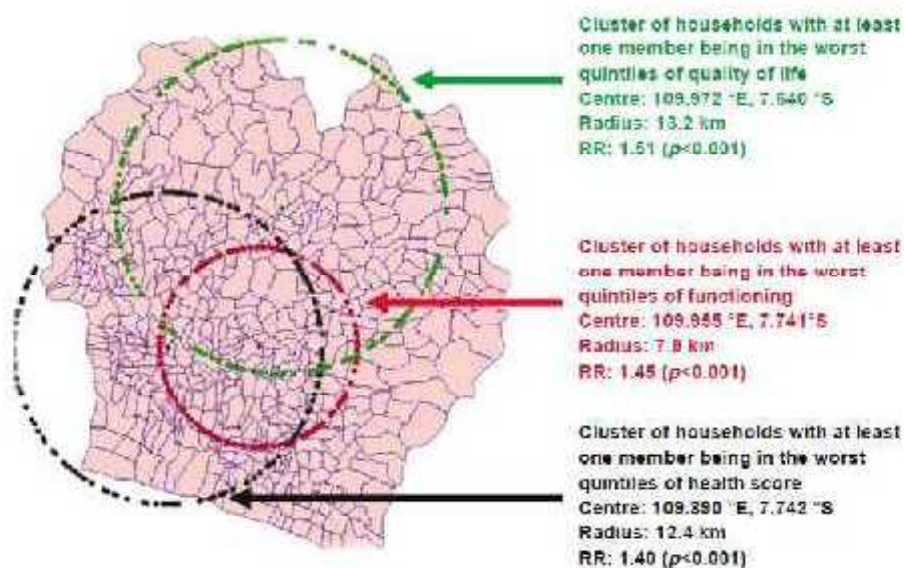


Fig. 2. Spatial distribution of poor health indices among 11,753 adults aged 50 years and over in Parworcio District, 2007.

Source: Ng et al (2010) Health inequalities among older men and women in Africa and Asia: evidence from eight Health and Demographic Surveillance System sites in the INDEPTH WHO-SAGE study. Global Health Action Supplement 2.





## Data analysis documentation

- Purpose - To ensure that the analyses can be properly reproduced
- Create a text file for all the relevant analyses
  - This text file needs to include both the relevant control file (with clear information about all the steps taken), as well as the output (with clear information for all results).
  - The text file needs to start off with the research question to be answered and the date of the analysis, and should end with a(n) (provisional) answer to the question
- Keep a copy of your data for the publication
- Keep a copy of all your programs



## Documenting results

- Almost every results section in an article starts with a paragraph about the recruitment of research participants.
- Many journals require a “patient flow chart” to be included in the article.
  - This represents how many patients were approached, which ones were selected and excluded (and the exclusion criteria), the dropouts and the number of patients ultimately remaining who participated

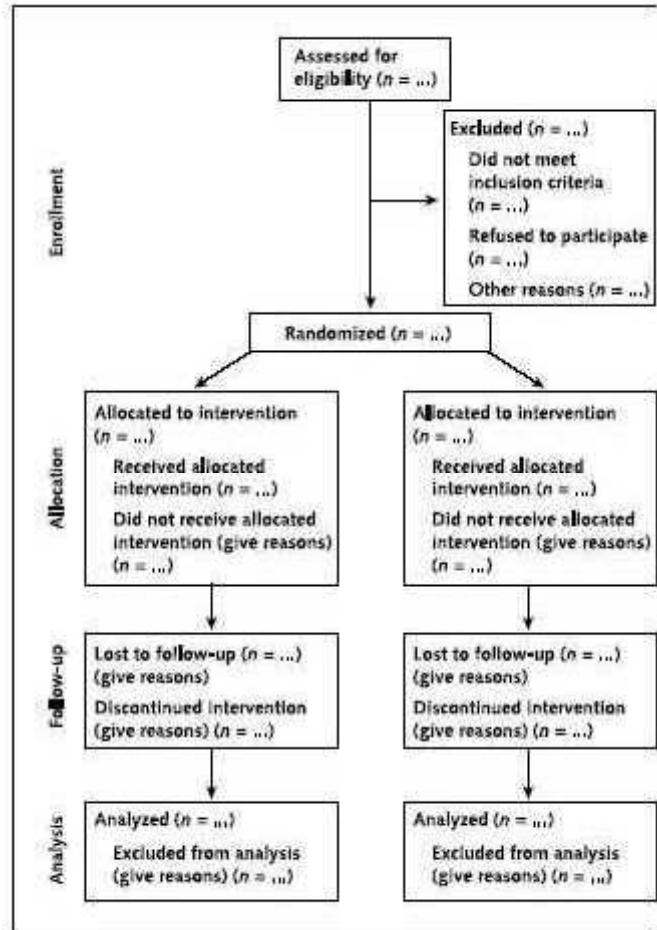






## Flow chart

Figure:  
Flow diagram of the process through  
the phases of a randomized trial  
(enrollment, intervention, allocation,  
follow-up and data analysis)



## Baseline table

- A baseline table (usually Table 1) is intended as a description of your research population.
  - Socio-demographic variables
  - clinical characteristics e.g. severity of the disorder; general health status
- All baseline values of the determinants, outcomes and potential prognostic variables should be included





## Reporting effects

- Effect estimates (e.g. average difference, relative risk or odds ratio) should always be included with the 95% confidence interval
- For (multiple) regression analysis the regression coefficient(s) (B) should be included for all cases, including the standard error(s) or a confidence interval.
- The p-value may also be included
- Advisable to include both the crude effect estimates, as well as any corrected effect estimates (e.g. corrected for age and gender).



### Example: study data

Table 4. Factors associated with poor self-rated health among 4,384 adults aged 50 and over in northern Ghana

Variables	Univariate model (OR and 95% CI)	Multivariate model (OR and 95% CI)
WILLDASL quintile		
Highest ability quintile	1.00	1.00
Second quintile	1.70 [1.43–2.22]**	1.65 [1.32–2.07]**
Third quintile	3.42 [2.72–4.23]**	3.18 [2.58–3.98]**
Fourth quintile	8.51 [5.16–8.17]**	5.75 [4.54–7.32]**
Lowest ability quintile	16.56 [10.1–20.6]**	14.20 [11.1–16.3]**
Sex		
Men	1.00	1.00
Women	1.54 [1.37–1.75]**	1.40 [1.28–1.73]**
Age group (year)		
50–59	1	1
60–69	1.52 [1.32–1.74]**	1.12 [0.96–1.31]
70 years and over	2.41 [2.00–2.82]**	1.24 [1.02–1.51]*
Education level		
Socioeconomic quintile		
Poorest quintile	1.00	1.00
Second quintile	1.17 [0.99–1.39]	1.12 [0.93–1.35]
Third quintile	1.05 [0.88–1.25]	0.95 [0.78–1.15]
Fourth quintile	0.88 [0.70–1.03]	0.77 [0.60–0.90]**
Least poor quintile	0.67 [0.52–0.87]**	0.65 [0.48–0.88]**

\*p < 0.05; \*\*p < 0.001.

Debpuur et al (2010) Self-reported health and functional status limitations among older people in the Kassena-Nankana District, Ghana. Global Health Action Supplement 2.





- Aim: To promote structured and targeted data analysis
- A stepwise plan created prior to the actual data analysis
- Intended as a starting point for the analysis
- **However, both the research questions and the analyses may be revised during the data analysis!**





## Research question

- The concrete research question needs to be formulated firstly within the analysis plan; this is the question intended to be answered by the analyses.
- PICO:
  - Population
  - Intervention
  - Comparison
  - Outcome(s)



## Which statistical techniques are to be used?

- Which (subgroup of the) population is to be included in the analyses?
- Data from which endpoint (T1, T2, etc) will be used?
- Which (dependent and independent) variables are to be used in the analyses and how are the variables supposed to be analysed (e.g. continuous or in categories)?
- Which variables are to be investigated as potential confounders of effect modifiers and how are these variables supposed to be analysed?
- Which analyses are to be carried out in which order (e.g. univariate analyses, multivariate analyses, analysis of confounders, analysis of interaction effects, analysis of sub-populations, etc.)?





## Example text

*“The overall proportion of current smokers, as well as the proportion by each relevant demographic and socioeconomic determinant, was calculated within each of the pooled datasets. Multivariable logistic regression was used to assess the association between current smoking and the potential demographic and socioeconomic determinants according to sex and country income group. All analyses were weighted accounting for the individual survey sample designs...Stata11 was used in all analyses.”*



## References/links

- Scientific style and format: the CBE manual for authors, editors, and publishers, 6th ed. Style Manual Committee, Council of Biology Editors. New York: Cambridge University Press, 1994
- Iverson C, Flanagin A, Fontanarosa PB, et al. American Medical Association manual of style: a guide for authors and editors, 9th ed. Hagerstown, Maryland, Lippincott Williams & Wilkins; 1997.





Are you ready to go now?



...what about your missing values...?



INTREC - INDEPTH Training & Research Centres of Excellence

# Statistical software and Missing values

an example







Now you know a lot about your data,  
potential analysis tools  
and about statistical software ...

please check, if your software does  
what you aiming for...

!!!



How does your statistic software deal with missing values?

- Frequency table of sex
- Mean age
- Dichotomize:  $\text{Age2} = (\text{Age} \leq 20)$
- Cross tabulation age \* sex
- How many observations remain for your regression analysis?

ID	Age	Sex	Weight
1	15	1	79.9
2	23	.	63.5
3	.	2	72.0





... any questions?



## Group discussion

### Tables and figures in SAGE publications

- Gomez-Olive et al (2010) Assessing health and well-being among older people in rural South Africa. *Global Health Action* Supplement 2.
- Van Minh et al (2010) Patterns of health status and quality of life among older people in rural Viet Nam. *Global Health Action* Supplement 2.

