



**Karolinska  
Institutet**

# Survey of data management and analysis within SIMSAM

Marie Reilly

*National SIMSAM steering group meeting 17 Sept 2012*

# Report from the SIMSAM mid-term evaluation 15-16<sup>th</sup> of November, 2011:



Karolinska  
Institutet

recommended sharing resources for data management, data documentation, data modeling and analysis

Stated that it as "unfortunate" that specific knowledge and expertise is not systematically shared

**.... "The panel strongly urges the SIMSAM nodes to find ways to do that as soon as possible"**

## **SIMSAM NATIONAL WORKSHOP, APRIL 2012**

# First conduct SIMSAM survey

- To learn what is being done already
- Identify potential for sharing resources
- Get suggestions for how to promote sharing
- Fuel discussion concerning how to proceed

# On-line questionnaire

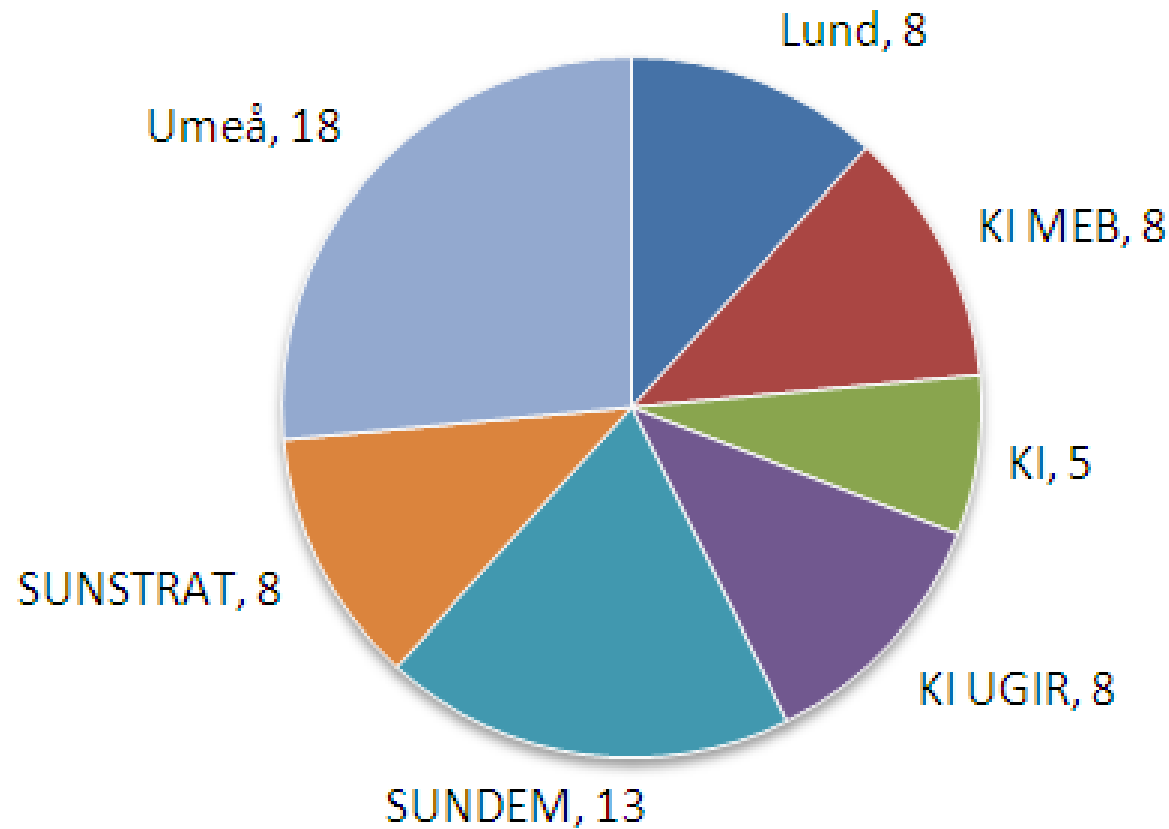
Input from April meeting and pilot testing

Sent out by Anna Holmström

Excellent response (n=72)

Preliminary results .....

# Responses from Nodes



# Role

PI/co-PI:	31%
Postdoc	19%
PhD	50%

# Main work in SIMSAM

Medical Science 50%

Social Science 47%

Epidemiologist 44%

Social Scientist 24%

Demographer 24%

Statistician 19%

Geographer 7%

Data Manager 7%

Database Designer

/Administrator 4%

Other (student, economist, clinician) 15%

# Data Preparation work

58% Yes

Stata	60%
SAS	28%
SPSS	23%
R	13%
ArcGIs	8%
Access	2.5%
dBase	2.5%



# Code for data preparation

**Found available code**      **31%**

class coding/origin, case/control, cohorts, families,  
standardised incidence.....

**Wrote code**      **87%**

case-control, cohorts, connecting kin, spouses/partners,  
siblings, identifying duplicates, ICD, class coding, person-  
time, time-dependent exposure

# Data Analysis work

87% Yes

Stata	53%
SPSS	27%
SAS	23%
R	18%
ArcGIs	6%
Other	16%

# Code for data analysis

Found available code	20/68	(29%)
Would recommend to others	10	(50%)
Needed to write code	23/62	(37%)
(relatively) easy for others		~ 50%

# Willing to share code

Within group	52/57 ( <b>91%</b> )
Within SIMSAM	42/52 ( <b>81%</b> )
Anyone	27/47 ( <b>57%</b> )

Interested in code from others: 55/72 (**76%**)

# Obstacles to sharing code

Lack of documentation/commenting	68%
Individual logic/style	65%
Finding resources	39%
Changing versions of software	36%

# Useful resources listed

Books, manuals, web pages, R package, SAS help,  
Lexis macro, stataлист, .....

# Conclusion

From survey, evidence of inefficiency and need for sharing, but also clear **opportunity**:

Interest

Willingness

Expertise

Information

# Acknowledgements

**Anna Holmström:** compiling email list, mailing

**Luwis Diya:** analysis of responses

**All respondents**



# Next steps?