



Combining and sharing of data

Catarina Almqvist Malmros. MD PhD
Professor and Consultant Pediatrician
Dept of Medical Epidemiology and Biostatistics (MEB) and
Astrid Lindgren Children's Hospital
PI SIMSAM-MEB family design



Combining and sharing of data

- Combining
 - register holders
 - researcher generated data
 - data cleaning
- Sharing
 - codes
 - legal and ethical issues
 - technical solutions

Combining

- SCB
 - Total population-, Multi-Generation and Migration registers
 - LISA (longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier) census (FoB)
 - Swedish Education-, Income and taxation-, Occupation registers
 - National school grade register (Year 9 grades Skolverket)
- National Board of Health and Welfare
 - Medical Birth register, Cause of Death register
 - National Patient register
 - Prescribed Drug register
 - Cancer register
- Swedish Twin Registry

Combining

- Stockholm county council
 - VAL, HAB
- Quality registers
 - Pastill, Bipolär, BUSA
 - RIKSÄT/Stepwise
 - Q-IVF, CF register, SNQ
 - National Diabetes Register and SWEDIABKIDS
 - CAH register (neonatal screening program)
- Dog ownership registers
 - Swedish Board of Agriculture National Dog Register
 - Swedish Kennel Club
 - Agria djurförsäkringar

Combining

- BRÅ / Crime registers
 - Lagföringsregistret
 - Misstankeregistret
 - Kriminalvårdsregistret
 - Rättsmedicinalverkets register
- Conscript register
- Universitetskanslerämbetet (UKÄ)
 - Swedish Scholastic Aptitude test (SweSAT), högskoleprovet
- Göteborg university pollen count laboratory
 - pollen counts from Göteborg and Bräkne-Hoby
- Lantmäteriet
 - air pollution (NOx, PM10)

Combining

- Researcher-generated data
 - Clinical cohorts
 - Patient data, medical records and biosamples
 - GWAS, EWAS, molecular epi data
- Legal, security and logistic issues
 - Virtual desktop infrastructure
 - Meta-analyses

Sharing – organisation within nodes

- development of the content and quality of registers
- RA/DBA coordinates the data work within the node(s)
 - introduce new members to data linkages or MONA
 - compile joint coding where this has been feasible
 - quick learning for newly hired and international group members
 - link between the nodes and Statistics Sweden and SoS
- node members have different areas of responsibilities
 - multigenerational register SCB, housing data, school data
 - National board of health and welfare data, quality registers
 - regular meetings on important issues such as structures for sharing code, cleaning and combining
 - RUT and registerforskning.se

Sharing – development of methods and software

- developed measures, methods and software valuable for register-based research, often available online
- SIMSEG
 - R-package for empirically calibrated agent based modeling
 - new method for estimating separate fixed effects for identical and fraternal twins;
 - methods for analysing influences on housing/neighbourhood choice using register data
 - R package bdynsys for bayesian dynamical systems modelling
 - new three-generations measures of ancestry
 - an interactive tool for visualizing trends in residential segregation

SUNDEM – combining data

- population-register systems of the Nordic region
 - similarity of Nordic societies as well as
 - similarity in infrastructures for research
- methods that enable joining of harmonised individual-level data from several countries without joining individual records of data
 - event-history models with aggregations of occurrences and exposure times -- derived from each country and then joined
- another project data linkage directly across Sweden and Finland
 - allows to follow all individuals who have ever moved between the two countries and who thus appear in both countries registers

SUNDEM – code sharing

- Micro-online system of Statistics Sweden (MONA).
 - national activities to stimulate code sharing among register researchers
 - not easy when applied to researchers who work with very different kinds of data
 - SUNDEM has developed an informal system of code sharing that also includes other demographers in Sweden (Lund, Umeå) and researchers at Stockholm University who use similar sources of data (SUDA, SOFI, Sociology, Human Geography)
 - code sharing is facilitated by the fact that most researchers do their programming in STATA

Additional projects to combine and share

- SCB Nordman project, to combine Nordic data
- MONA system (Microdata Online Access)
 - SCB
 - good technical support
 - excludes SoS and other registers

SIMSAM Lund

- Anonymised data extractions have been used for educational purposes;
 - Malmö Diet and Cancer Cohort (Lund)
 - Scania Public Health Survey (Lund)
 - MFR + military conscription data on brothers (Lund + KI)
 - Case-control data on MS (Lund + KI)
- Research education courses in epidemiology and register-based research at Lund university
 - lectures, computer labs and projects
 - replication of published results
 - Complementary sensitivity analyses

GOCARTs

- Publications on combined data from NDR, Patient- and Cause of Death registers and the total population register (controls).
 - Ståhl CH, Lind M, Svensson AM, Gudbjörnsdóttir S, Mårtensson A, Rosengren A. Glycaemic control and excess risk of ischaemic and haemorrhagic stroke in patients with type 1 diabetes: a cohort study of 33 453 patients. *J Intern Med.* 2017 Mar;281(3):261-272.
 - Hedén Ståhl C, Lind M, Svensson AM, Kosiborod M, Gudbjörnsdóttir S, Pivodic A, Clements M, Rosengren A. Long-term excess risk of stroke in people with Type 2 diabetes in Sweden according to blood pressure level: a population-based case-control study. *Diabet Med.* 2017 Apr;34(4):522-530

SIMSAM-MEB family design

- Family design
 - Twins
 - Siblings and half siblings
 - Cousins and half cousins
- to account for confounding from genetic and early environmental factors

JAMA | Original Investigation

Associations of Maternal Antidepressant Use During the First Trimester of Pregnancy With Preterm Birth, Small for Gestational Age, Autism Spectrum Disorder, and Attention-Deficit/Hyperactivity Disorder in Offspring

Ayesha C. Sujan, MA; Martin E. Rickert, PhD; A. Sara Öberg, MD, PhD; Patrick D. Quinn, PhD; Sonia Hernández-Díaz, MD, PhD; Catarina Almqvist, MD, PhD; Paul Lichtenstein, PhD; Henrik Larsson, PhD; Brian M. D'Onofrio, PhD

JAMA. 2017;317(15):1553-62
doi: 10.001/jama.2017.3413.

LETTERS

Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology

Manuel A Ferreira^{1,3,4}, Judith M Vonk^{2,3,4}, Hansjörg Baurecht^{1,3,4}, Ingo Marenholz^{4,5,38}, Chao Tian⁶⁻³⁸, Joshua D Hoffman^{7,38}, Quanta Helmer^{8,38}, Annika Tillander^{9,38}, Vilhelmina Uilemar^{9,38}, Jenny van Dongen^{9,38}, Yi Lu¹⁰, Franz Rischendorf^{11,38}, Jorge Esparza-Gordillo^{12,37}, Chris W Medway¹³, Edward Moutonjoy¹⁴, Kimberley Burrows¹⁵, Oliver Hummel¹⁶, Sarah Grosche¹⁷, Ben M Brampton¹⁸⁻²¹, John S Witte²², Jonke-Jan Hottenga²³, Gonneke Willemssen²⁴, Jie Zheng²⁵, Elke Rodriguez²⁶, Melanie Hotze²⁷, Andre Franke¹⁴, Joana A Revez¹, Jonathan Beesley¹, Melanie C Matheson¹⁷, Shyamali C Dharmage¹⁷, Lisa M Bain¹, Lars G Fritsche¹¹, Maiken E Gabrielsen¹¹, Brunilda Balliu¹⁶, 23andMe Research Team¹⁷, AAGC collaborators¹⁷, BIOS consortium¹⁷, LifeLines Cohort Study¹⁷, Jonas B Nielsen¹⁸⁻¹⁹, Wei Zhou¹⁹, Kristian Hveem^{11,20}, Arnefjell Langhammer²⁴, Oddgeir I. Holmen¹, Mari Loev^{11,21}, Gonçalo R Abecasis^{11,22}, Cristen J Willer^{11,18,19,23}, Andreas Arnold²³, Georg Homuth²⁴, Carsten O Schmidt²⁵, Philip J Thompson²⁶, Nicholas G Martin¹, David I Duffy¹, Natalija Novak²⁷, Holger Schulz^{28,29}, Stefan Karrasch²⁸⁻³⁰, Christian Gieger³¹, Konstantin Strachan³², Ronald B Melle³³, David A Hinds⁶, Norbert Hübner^{4,39}, Stephan Weidinger^{4,39}, Patrik K E Magnusson⁴⁰, Rick Jansen^{41,39}, Eric Jorgenson^{42,39}, Young-Ae Lee^{43,39}, Dorret I Boomsma^{44,39}, Catarina Almqvist^{45,39}, Robert Karilsson⁴⁶, Gerard H Koppelman^{47,39} & Lavinia Paternoster^{48,39}

Nat Genet. 2017;49(12):1752-7. doi: 10.038/ng.3985

Catarina Almqvist Malmros 16 May 2018 15

JAMA | Original Investigation


Associations of Maternal Antidepressant Use During the First Trimester of Pregnancy With Preterm Birth, Small for Gestational Age, Autism Spectrum Disorder, and Attention-Deficit/Hyperactivity Disorder in Offspring

Ayesha C. Sujan, MA; Martin E. Rickert, PhD; A. Sara Öberg, MD, PhD; Patrick D. Quinn, PhD; Sonia Hernández-Díaz, MD, PhD; Catarina Almqvist, MD, PhD; Paul Lichtenstein, PhD; Henrik Larsson, PhD; Brian M. D'Onofrio, PhD

JAMA. 2017;317(15):1553-62
doi: 10.001/jama.2017.3413.

- Children born 1996-2013 MedicatI birth reg N=1,5 million births
- Family design to account for confounding (genetic and early environmental)
 - siblings, half siblings, cousins, half siblings fathers through MGR
- Exposures from the Prescribed Drug register and MBR
- Outcomes from the PDR and National Patient register
- Analyses before, during and after 1st trimester
- Family design on siblings and cousins discordant for E and O
- Paternal negative control

Catarina Almqvist Malmros 16 May 2018 16



**Karolinska
Institutet**

- Based on Swedish Twin Registry Child and Adolescent Twin Study in Sweden cohort
- Asthma, hay fever and eczema
 - parental-reported questionnaire data
 - register-based from Prescribed Drug- and National Patient register
- GWAS data
- Meta-analyses
- N=360,000 children

LETTERS


**nature
genetics**

Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology

Manuel A Ferreira^{1,3,4}, Judith M Vonk^{2,3,8}, Hansjörg Baurecht^{1,3,8}, Ingo Marenholz^{4,5,38}, Chao Tian^{6,38}, Joshua D Hoffman^{7,38}, Quanta Helmer^{9,38}, Annika Tillander^{9,38}, Vilhelmina Uilemar^{9,38}, Jenny van Dongen^{9,38}, Yi Lu^{9,38}, Franz Rüschenhoff^{9,38}, Jorge Esparza-Gordillo^{4,5,37}, Chris W Medway¹⁰, Edward Mounjoy¹⁴, Kimberley Burrows¹⁰, Oliver Hummel¹, Sarah Grosche^{4,5}, Ben M Brampton^{10,32}, John S Witte¹¹, Jonke-Jan Hottenga⁸, Gonneke Willemssen⁸, Jie Zheng¹⁰, Elke Rodriguez³, Melanie Hotze³, Andre Franke¹⁴, Joana A Revez¹, Jonathan Beesley¹, Melanie C Matheson¹², Shyamali C Dharmage¹³, Lisa M Bain¹, Lars G Fritsche¹¹, Maiken E Gabrielsen¹¹, Brunilda Balliu¹⁶, 23andMe Research Team¹⁷, AAGC collaborators¹⁷, BIOS consortium¹⁷, LifeLines Cohort Study¹⁷, Jonas B Nielsen^{16,19}, Wei Zhou¹⁹, Kristian Hveem^{11,20}, Arneff Langhammer²¹, Oddgeir I. Holmen¹, Mari Loevdal^{11,21}, Gonçalo R Abecasis^{11,22}, Cristen J Willer^{11,18,19,23}, Andreas Arnold²³, Georg Homuth²⁴, Carsten O Schmidt²⁵, Philip J Thompson²⁶, Nicholas G Martin¹, David I Duffy¹, Natalija Novak²⁷, Holger Schulz^{28,29}, Stefan Karrasch^{28,30}, Christian Gieger³¹, Konstantin Strachan³², Ronald B Meles³³, David A Hinds⁶, Norbert Hübner^{4,39}, Stephan Weidinger^{4,39}, Patrik K E Magnusson²⁹, Rick Jansen^{4,39}, Eric Jorgenson^{4,39}, Young-Ae Lee^{4,39}, Dorret I Boomsma^{8,39}, Catarina Almqvist^{4,39}, Robert Karilsson^{4,39}, Gerard H Koppelman^{4,39} & Lavinia Paternoster^{10,39}

Nat Genet. 2017;49(12):1752-7. doi: 10.038/ng.3985

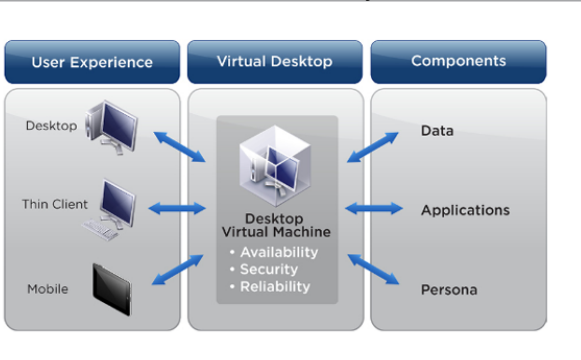
Catarina Almqvist Malmros
16 May 2018
17



**Karolinska
Institutet**

SIMSAM-MEB family design

- Virtual Desktop Infrastructure
- Access to MEBs IT-resources remotely in a secure environment



Catarina Almqvist Malmros
16 May 2018
18

Combining and sharing

- Combining
 - register holders
 - researcher generated data
 - data cleaning
 - knowledge dissemination
- Sharing
 - codes
 - legal, security and ethical issues
 - technical solutions

Thank you!