



Data Resource Profile

Data Resource Profile: Swedish Microdata Research from Childhood into Lifelong Health and Welfare (Umeå SIMSAM Lab)

Urban Lindgren,^{1*} Karina Nilsson,² Xavier de Luna³ and Anneli Ivarsson⁴

¹Department of Geography and Economic History, ²Department of Sociology, ³Department of Statistics, Umeå School of Business and Economics, ⁴Epidemiology and Global Health Unit, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

*Corresponding author. Department of Geography and Economic History, Umeå University, SE-901 87 Umeå, Sweden.

E-mail: urban.lindgren@umu.se

Accepted 9 December 2015

Data resource basics

The childhood period as key to future health and welfare

In order to ensure a good start to life for every child there is a need for research on social determinants of health and, in wider social and economic spheres, to achieve greater health equity.¹ Different sorting processes lead to people entering upon various paths early on in life and, as an echo from the past, life opportunities are shaped by previous and current family circumstances, alongside people's aspirations, wishes and abilities (Figure 1). Failing to make necessary investments in children and adolescents based on their overall needs is neither economically nor socially sustainable.²

Sweden has a 'gold-mine' of observational data

Research on childhood and its influence on lifelong health and welfare is facilitated by individual-level longitudinal data from the cradle to the grave, and data on the family and other social contexts that the person is part of during different periods of life. In Sweden, the potential for such research is outstanding since individual-level data are available through a large number of registers, and linkage is possible by means of personal identity numbers.^{3,4} The personal identity number is unique for each person, and is

essential for living in Swedish society, e.g. to have a bank account, to access health care and other social services. These personal identity numbers are used by authorities in all administrative and health registers. For research purposes, and after approval from the Ethical Vetting Board and register holders, data sources can be linked at the individual level through this unique number.

The long tradition of collecting data for various population registers makes it possible to create long time-series of repeated measures for each individual. Demographic, socio-economic and health data on the entire population can be combined in different ways as a basis for interdisciplinary projects posing novel research questions in the nexus between the social and medical sciences.

Swedish Initiative for Microdata Research in the Social and Medical Sciences

In 2008, the Swedish Research Council launched the research called 'Swedish Initiative for Research on Microdata in the Social and Medical Sciences' (SIMSAM), which aimed at strengthening register data research and supporting a new generation of scholars within this field.^{5,6} The term microdata refers to individual-level data in registers and databases. Our interdisciplinary research group was awarded a grant that made it possible to start up a research

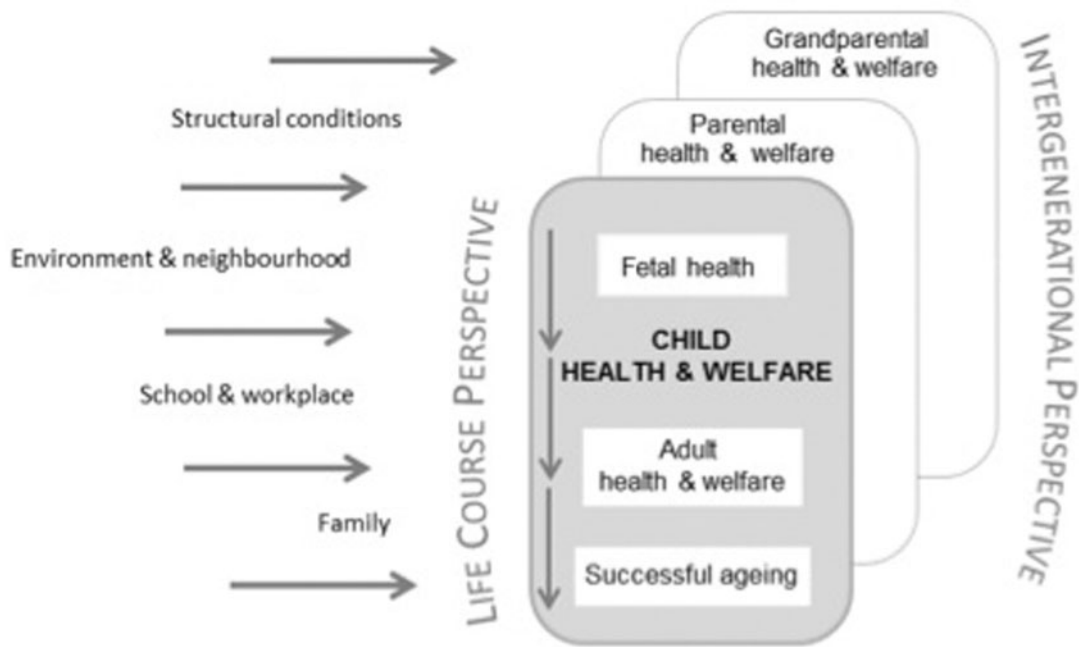


Figure 1. The Umeå SIMSAM Lab conceptual model for a life-course approach to health and welfare considering individuals' life course, the intergenerational perspective and different levels of influence from the family, neighbourhood etc.

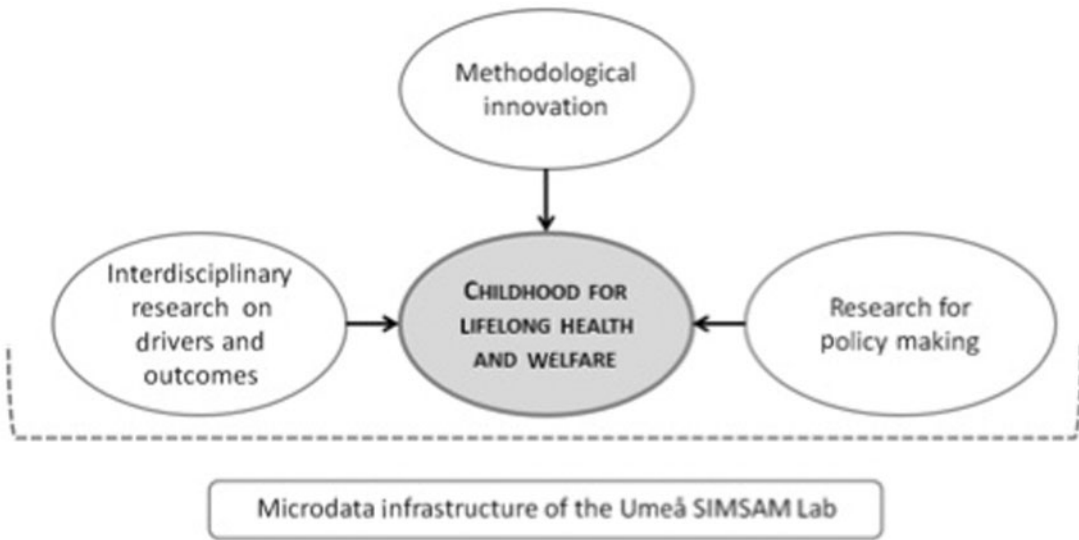


Figure 2. The Umeå SIMSAM Lab main research fields.

infrastructure, the Umeå SIMSAM Lab, described in this profile. Our research agenda is illustrated in Figure 2. Based on the data infrastructure, we are developing a creative and multidisciplinary research environment, including seminars to which national and international researchers from different disciplines contribute.

The Umeå SIMSAM Lab data infrastructure offers unique research opportunities

The Umeå SIMSAM Lab data has been compiled from several national and regional register and database sources, all hosting individual-level prospectively collected data based on each person's unique identity number. On behalf of our



Figure 3. The Umeå SIMSAM Lab data cover the whole population of Sweden, with enriched data on Västerbotten region’s population.

Table 1. The prospective and longitudinal character of the Umeå SIMSAM Lab data with measures on all 14.9 million individuals in Sweden from 1960 to 2010

| <div>Year</div> <div>Characteristics</div> | 1960 | 1970 | 1980 | 1990 | 2000 | 2010 |
|--|-------------|-------------|-------------|-------------|-------------|-------------|
| <div>NATIONAL DEMOGRAPHIC AND SOCIOECONOMIC MEASURES</div> <div><ul style="list-style-type: none">Censuses (demographic & socio-economic)DemographySocio-economics</div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |
| <div>NATIONAL HEALTH MEASURES</div> <div><ul style="list-style-type: none">Medical Birth RegisterNational Patient RegisterCeliac Disease RegisterPrescribed Drug Register</div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |
| <div>REGIONAL MEASURES</div> <div><ul style="list-style-type: none">Västerbotten Intervention ProgrammeInjury DatabaseSalut Programme</div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |
| <div>GEOGRAPHICAL COORDINATES</div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |
| <div>INTERGENERATIONAL LINKS</div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |

research group, Statistics Sweden⁷ has combined data from the different sources on each individual, and in this process data were anonymized by replacing the personal identity number with an internal serial number.

The Umeå SIMSAM Lab hosts longitudinal register data covering demographic, socioeconomic and health indicators for the entire Swedish population over several decades, complemented by geographical residential coordinates and

Table 2. The Umeå SIMSAM Lab microdata sources and characteristics

| REGISTER HOLDERS | SCB ^a | SoS ^b | UmU ^c | Västerbotten County Council | | |
|----------------------------------|---|--|--|---|---|--|
| Registers and databases | Total population Censuses Geography Multi-generational Pupil register University register Firm register | Medical Birth Register (<i>MFR</i>) National Patient Register (<i>NPR</i>) Prescribed Drug Register (<i>PDR</i>) | National Celiac Disease Register | Västerbotten Intervention Programme | Injury Database | Salut Programme ^d |
| Links on individual level | Personal identity number | → | → | → | → | → |
| Spatial resolution | Urban areas: 250*250 m Rural areas: 1*1 km | → | → | → | → | → |
| Geographical extension | Sweden | Sweden | Sweden | Västerbotten County | Umeå Region | Västerbotten County |
| Time coverage | Censuses: 1960-90 Annually 1968-2010 | <i>MFR</i> : 1973-2010 <i>NPR</i> : 1987-2010 <i>PDR</i> : 2005-10 | 1991-2010 | 1990-2010 | 1998-2010 | 2008-10 |
| Time resolution | Census: 5-years Registers: Daily/yearly | Daily | Monthly | Daily | Daily | Daily |
| Ages | All ages | All ages | 0-14.9 years | 40, 50 and 60 years | All ages | Parents-to-be and 18 months-olds |
| Variable examples | Demography Family links Country of origin Residence Education School grades Work Income Social benefits Coordinates: residence and workplace | Mother's weight/height Date/mode of delivery Gestational age APGAR-scores ^e Child's birth weight/length Date of hospital admission/discharge, and diagnosis (ICD) ^f Prescription drugs with respect to amount, date, and type (ATC) ^g | Symptomatic or screening-detected Mode for diagnosis Month/year of diagnosis Hospital | Self-rated health Health related quality of life Lifestyle habits Social support and networks Weight & height (BMI) Blood pressure Blood lipids Serum-glucose when fasting | Type of injury Type of trauma Circumstances around the trauma Treatment Diagnosis | Parents' education and work Parents' self-rated health Parental-rated child health Social support and networks On parents' and the children: Weight & height (BMI) Lifestyle habits |

^aSCB: Statistics Sweden.^bSoS: National Board of Health and Welfare.^cUmU: Umeå University.^dSalut Child Health Promoting Intervention Programme.^eAPGAR: reflecting the health of the newborn.^fICD: International Classification of Diseases.^gATC: Anatomic Therapeutic Chemical classification system.

intergenerational links. This provides an opportunity to follow individuals along their life course in order to understand, for example, mechanisms triggering social problems and ill health. Thus, it is possible to link life situations during childhood and adolescence to adult life outcomes in different social domains. Moreover, for each individual register, data at hand provide information about individual-specific sets of other people and how these sets change with respect to demographic, socioeconomic and health-related

characteristics as the individual grows older. The relational characteristics of the data thus offer information about kinship and potential social exchange that may take place in residential neighbourhoods, schools and workplaces.

Data resource area and population coverage

The Umeå SIMSAM Lab data resource covers the entire Swedish population during the period 1960 to 2010. The

number of individuals included amounts to 14.9 million, i.e. all people registered as living in Sweden at some point in time during the period. Furthermore, a large number of parameters including demographics, socioeconomic and medical/health data are available. Our data are geo-referenced on a detailed spatial scale (hectare grid) annually indicating the place of residence and workplace of all individuals included. Even richer information is available at some regional levels by systematic and long-term initiatives within the health care system involving data collection as illustrated in Figure 3.

Time coverage and frequency of measurements

The time coverage of the measured variables depends on the sources of information, i.e. from which register or programme the data were obtained. The different windows of time coverage are illustrated in Table 1.

Census data (demographic, geographical and socioeconomic data) are available every 5 years from 1960 to 1990.⁷ Register data on demography and geography are available annually since 1968, with daily time stamps for life events including birth, civil status changes and deaths, and more detailed socioeconomic register data have been available annually from 1985 onwards.⁷ The National Medical and Health registers were established at different dates, thereby yielding different time coverages, and are available with a daily time stamp.⁸ The Medical Birth Register started in 1973, the National Hospitalisation Register in 1987, the National Celiac Disease Register in 1991 and the Prescribed Drug Register in 2005.

The regional data sources (with daily time stamps) have different time spans and frequencies. The Västerbotten Intervention Programme is a health promotion programme for the middle-aged people of the Västerbotten County (northern Sweden), initiated in 1985 and with data available from 1990 onwards.⁹ Another Västerbotten programme from which data have been obtained is the Salut Programme, which is a child health-promoting intervention programme that was initiated in 2005 and has yielded data from 2008.¹⁰ Finally, at a more local level, we have information from 1998 on all injuries resulting in hospitalization or out-patient specialized care.¹¹ The latter data cover individuals living within the catchment area of the University Hospital of Northern Sweden in Umeå.

Measures

All individuals included in the Umeå SIMSAM Lab infrastructure have data on many variables from various domains. Note that in the national health registers, health is

largely measured through its opposite, i.e. disease data, whereas data on the population in the Västerbotten region have been enhanced by self-rated health, lifestyle measures, etc. (Figure 3). The data sources and the prospectively collected measures are described below and summarized in Table 2. An exhaustive description of the data can be obtained from the authors.

Demographic and socioeconomic measures for the entire Swedish population

In the Umeå SIMSAM Lab there are demographic and socioeconomic measures on each individual covering several decades with, for most of them, repeated measures over the life course. These measures have been made available by Statistics Sweden.⁷ Examples of measures are: birth date, sex, place of residence, grades in compulsory school, secondary education (type of education and grades), highest achieved education, civil status, employment status, taxable income, transfer payments related to sick leave, childbirth, parental leave and retirement pension. Most nationally collected administrative measures concern adults, but intergenerational links make it possible to view children indirectly through their parents.

Health variables for the entire Swedish population

Health and/or medical data are available for almost every individual living in Sweden from 1973 up to 2010, with most of them having repeated measures over the life course. These measures have been made available by the National Board of Health and Welfare⁸ to which health authorities and pharmacies across the country are obliged to report.

Through the Medical Birth Register there are measures on all infants and their mothers. Examples of measures are: maternal smoking, mother's weight and height at first visit to antenatal care and at delivery, maternal health problems, date of delivery, gestational age at delivery, mode of delivery, APGAR scores, child's birthweight and length and any diagnosis related to pregnancy or to the newborn. The National Patient Register contributes with measures on all in-hospital care with respect to date of admission and discharge, and diagnosis using the International Classification of Diseases (ICD) code system. The Prescribed Drug Register provides individual measures with regard to all prescription drugs that have been supplied by pharmacies to individual patients with respect to amount, date and type according to the Anatomic Therapeutic Chemical classification system (ACT code).

In addition, the Lab includes measures from the National Swedish Childhood Celiac Disease Register, which is the only one of its kind worldwide.^{12,13} Examples

of measures are: symptoms before diagnosis and treatment, initial serological marker results, small intestinal biopsy pathological-anatomical evaluations and how reliable the individual celiac disease diagnosis is judged to be.

Regional demographic, socioeconomic, health and lifestyle variables

The Västerbotten Intervention Programme (VIP) combines epidemiological surveillance and health-promoting interventions⁹ and is run by the Västerbotten County Council, which is the health authority of the region. Everyone in the region is invited when turning 40, 50 and 60 years. Up to 2010 the VIP database encompassed 103 000 unique individuals, and out of these circa 45 000 had taken part more than once. Many of the individuals captured in the database play an important role in children's lives as parents or grandparents. Data collection takes place by means of a comprehensive questionnaire (e.g. self-rated health and lifestyle habits), measured health data (e.g. blood pressure, weight, height and waist circumference) and analyses of blood samples (e.g. blood lipids and serum glucose when fasting and 2 h after a glucose load). However, Umeå SIMSAM Lab does not have access to the blood samples, thus additional analyses are not possible.

The Umeå Injury Database is developed according to the European Home and Leisure Accident Surveillance System.¹¹ The surveillance encompasses 150 000 injuries, of which 45 000 concern children. When entering the University Hospital of Northern Sweden with any injury, the patient is requested to fill in a questionnaire and later diagnosis and treatment information is retrieved from the medical record. Examples of measures are: type of injury and trauma, including information about where the injury took place and alcohol consumption.

The Salut Child Health Promoting Intervention Programme is run by the Västerbotten County Council. It combines epidemiological surveillance with systematic multi-sector health-promoting interventions.¹⁰ The Programme targets all parents-to-be and all children (0–19 years) in the region. The Umeå SIMSAM Lab data are so far limited to parents-to-be and 18-month-olds and their parents. Examples of measures are weight and height, lifestyle habits, experience of social support, adult self-rated health and parent-rated child health.

Geographical coordinates and intergenerational links

All individuals have intergenerational links to other individuals, e.g. parents and siblings, cousins and their parents, and grandparents. In addition, each individual is characterized

by geographical residential and workplace information. These relational characteristics bring additional possibilities by offering information about kinship and potential social exchange that may take place in residential neighbourhoods, schools and workplaces.

Data resource use

Within a field that benefits from alliances between disciplines, the formation of the Umeå SIMSAM Lab has sparked a range of interdisciplinary innovative collaborations and research projects. The Umeå SIMSAM Lab data resources were made available for researchers in mid 2014, and our work in forming a new platform for interdisciplinary research has provided us with teams of researchers running a wide palette of projects, of which a few examples are presented below.

Example 1: Family organization and overweight

Overweight has become a growing public health issue in many parts of the world, especially in high-income countries where overweight and obesity prevalence has increased significantly.¹⁴ Numerous studies have investigated the influence of family organization on childhood obesity, looking at factors like divorce and separation,¹⁵ marriage vs cohabitation,¹⁶ residential instability (e.g. moving between parents' homes)¹⁶ and living with one or both parents.¹⁷ Less is, however, known about the long-term impact of family composition and obesity in adulthood. We are therefore currently studying the effects of childhood family history on adult overweight and obesity. In particular, we are able to include a broad range of family factors: socioeconomic and health circumstances, changes in family composition and family disruptions.

Example 2: Celiac disease causes and consequences

Celiac disease was previously a rare disease of European children, but has over recent decades emerged as global public health challenge involving all ages.^{12,13,18} We have earlier explored causes behind celiac disease with the ultimate goal of moving towards preventive strategies.^{12,13,19,20} Our findings suggested that infant feeding practices contribute,^{12,19} and that the local context is an important explanatory level.²⁰ Currently, the Umeå SIMSAM Lab is allowing us to further explore potential risk factors for the disease (e.g. family socioeconomic and perinatal conditions), and to study long-term consequences of living with the disease (e.g. with respect to both health and school grades).

Example 3: Methodological innovation for register-based research

The combination of advanced scientific queries and multidimensional and complex data calls for methodological innovation. For instance, a major issue in the design of observational studies conducted at the Umeå SIMSAM Lab is the multidimensionality of the data, which stems from both millions of individuals and thousands of characteristics, some inherited from individuals sharing domains.^{21,22} Although large samples yield a drastic increase in precision, the latter may critically depend on assumed models. Thus, we are developing relevant non-parametric study designs and methods for sensitivity analyses based on previous work.^{23–26}

Strengths and weaknesses

The comprehensive set of register data available at the Umeå SIMSAM Lab offers a wide variety of opportunities. However, comprehensive data covering the population of an entire nation over a timespan of 50 years inevitably calls for caution and a need to be knowledgeable about changes over time with respect to data sources available, data collection methods, variable definitions, quality aspects etc. Register data produced by Swedish authorities are commonly regarded reliable measures.

Accessing the complete population enables analysis of various rare events and conditions that often are too infrequent to account for in smaller data sets. Furthermore, since data are longitudinal it is possible to follow individuals over long periods of time, thus enabling life-course analyses and combining observations from different phases of life. The Umeå SIMSAM Lab database includes links between family members, which make it possible to account for family background and parental characteristics. As the multigenerational register allows linking individuals with the extended family, it is possible to account for earlier generations (i.e. grandparents) as well as relatives (e.g. aunts and uncles). Another strength is that data are geo-referenced on a detailed scale, which makes it possible to characterize neighbourhoods of various sizes, thus opening up possibilities to study different forms of residential segregation, or various consequences of having to travel long distances to other family members, schools or health care services.

There are of course limitations to our data resources. All data sources are not available for the entire time period covered (Table 1). The earliest health data provided are those of the Medical Birth Register that started in 1973. Moreover, the demographic and socioeconomic data are not homogeneous over the period, with the period from 1985 to 2010 containing richer information on, for example, income and family forms than the earlier period.

Data resource access

The Umeå SIMSAM Lab welcomes researchers to make contact for collaborative projects and apply for access to the data resource. Interested researchers are encouraged to contact the Umeå SIMSAM Lab steering group or researchers already conducting studies using this data resource. To access data, research groups need to fill in an application form stating the study objectives, data needs, etc. The proposed project needs to relate to the overall research programme of the Umeå SIMSAM Lab, as outlined earlier in this article. The application will be administered by the steering group and, if approved, researchers are offered access to the data and state-of the art hardware and statistical software, as well as access to qualified support from data managers and technicians. Access to data is restricted to a secure data Lab, in accordance to the laws

The Umeå SIMSAM Lab in a nutshell

- The Lab was established for interdisciplinary research from childhood into lifelong health and welfare. It provides individual-level data on an entire nation's population, covering demography, socioeconomics and health, otherwise rarely accessible to the research community.
- Initiated in 2008 and launched in 2014, the Lab hosts comprehensive prospectively collected individual-level data on all 14.9 million individuals in Sweden between 1960 and 2010, including intergenerational links and geographical coordinates. The foundation is data from national registers at Statistics Sweden, rich in demography and socioeconomics. Health data come from national registers such as the Medical Birth Register, the Patient Register, the Prescribed Drug Register and the Celiac Disease Register. Regionally data have been enriched by systematic and long-term initiatives involving data collection on weight/height, lifestyle habits, self-rated health, injuries etc. Each person has a unique serial number that allows linkage on individual level from all data sources.
- The combination of longitudinal and individual-level demographic, socioeconomic and health data provides unprecedented research opportunities. In addition, social relations within family, workplace and neighbourhood can be depicted as intergenerational links, and geographical coordinates are available.
- Researchers are welcomed to make contact for collaborative projects. Contact information and details on data access are available on: [www.simsam.org.umu.se].

and regulations surrounding individual data at this level of detail. Data cannot be removed from the Lab, and it is of vital importance that researchers observe security routines.

Contact information and details of how to access data are available on our website: [www.simsam.org.umu.se]. Under 'Data access' there is a policy document describing routines on how to apply for data access, data retrieval, use of the data etc.

Funding

This work was supported by the Swedish Research Council [grant number 2008-7491] and by strategic financial support from Umeå University.

Acknowledgements

We are grateful to Lena Mustonen at the Epidemiology and Global Health Unit at Umeå University for excellent continuing administrative support, and for help in finalizing this article. Key persons for facilitating our efforts within Umeå SIMSAM Lab are Jenny Häggström and Erik Bäckström, as database manager and data technician, respectively. Sincere thanks go to the steering groups of registers and databases that share our vision of increased usefulness by combining national and regional population data: the Västerbotten Intervention Programme Database, the National Swedish Childhood Celiac Disease Register, the Umeå Injury Database and the Salut Programme Database.

Conflict of interest: None declared

References

- Marmot M, Friel S, Bell R, Houweling TA, Taylor S; Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 2008;**372**:1661–69.
- Irwin LG, Siddiqi A, Hertzman C. *Early Child Development: A Powerful Equalizer*. Final Report for the World Health Organization's Commission on the Social Determinants of Health. Vancouver, BC: Human Early Learning Partnership (HELP), 2007.
- Rosén M. Översyn av de nationella kvalitetsregistren. *Guldgruvan i hälso- och sjukvården. Förslag till gemensamsatsning 2011–2015. [Revision of The National Quality Registers. The Goldmine in Health Care. Proposal for a Joint Investment 2011–2015.]* Stockholm: LjungbergsTryckeri, 2010.
- Unik kunskap genom registerforskning. *Betänkande av Registerforsknings-utredningen. [Unique Knowledge Through Register-based Research. A Report by the Register-based Research Investigation.]* Statens Offentliga Utredningar, SOU 2014:45. Stockholm; Elanders Sverige AB, 2014.
- SIMSAM: Swedish Initiative for Research on Microdata in the Social and Medical Sciences. <http://www.simsam.nu> (6 July 2015, date last accessed).
- Stenbeck M, Andersson G, Berglund A *et al.* (eds). *Swedish Registers. A Unique Resource For Health And Welfare*. Stockholm: Vetenskapsrådet, 2013.
- Statistics Sweden. Våra tjänster. <http://www.scb.se/sv/Vara-tjanster/SCBs-data-for-forskning/SCBs-datalager/> (6 July 2015, date last accessed).
- The National Board of Health and Welfare/Health Data. <http://www.socialstyrelsen.se/register/halsodataregister/> (6 July 2015, date last accessed).
- Norberg M, Wall S, Boman K, Weinehall L. The Västerbotten Intervention Programme: background, design and implications. *Glob Health Action* 2012;**5**:1–9.
- Eurenius E, Lindkvist M, Sundqvist M, Ivarsson A, Mogren I. Maternal and paternal self-rated health and BMI in relation to lifestyle in early pregnancy: The Salut Programme in Sweden. *Scand J Public Health* 2011;**39**:730–41.
- Röding F, Lindkvist M, Bergström U, Lysholm J. Epidemiologic patterns of injuries treated at the emergency department of a Swedish medical center. *Inj Epidemiol* 2015;**2**:3:1–8.
- Ivarsson A, Persson LÅ, Nyström L *et al.* Epidemic of coeliac disease in Swedish children. *Acta Paediatr* 2000;**89**:165–71.
- Namatovu F, Sandström O, Olsson C, Lindkvist M, Ivarsson A. Celiac disease risk varies between birth cohorts, generating hypotheses about causality: evidence from 36 years of population-based follow-up. *BMC Gastroenterol* 2014;**14**:59.
- Chaparro P, Ivarsson A, Koupil I *et al.* Regional inequalities in pre-pregnancy overweight and obesity in Sweden, 1992, 2000, and 2010. *Scand J Public Health* 2015;**43**:534–39.
- Schmeer KK. Family structure and obesity in early childhood. *Soc Sci Res* 2012;**41**:820–32.
- Bzostek SH, Beck AN. Familial instability and young children's physical health. *Soc Sci Med* 2011;**73**:282–92.
- Chen AY, Escarce JJ. Family structure and childhood obesity: An analysis through 8th grade. *Matern Child Health J* 2014;**18**:1772–77.
- Green PH, Lebwohl B, Greywoode R. Celiac disease. *J Allergy Clin Immunol* 2015;**135**:1099–106.
- Ivarsson A, Myleus A, Norstrom F *et al.* Prevalence of Childhood Celiac Disease and Changes in Infant Feeding. *Pediatrics* 2013;**131**:E687–94.
- Namatovu F, Strömgen M, Ivarsson A *et al.* Neighborhood conditions and celiac disease risk among children in Sweden. *Scand J Public Health* 2014;**42**:572–80.
- de Luna X, Waernbaum I, Richardson TS. Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* 2011;**98**:861–75.
- Persson E, Häggström J, Waernbaum I, de Luna X. Data-driven algorithms for dimension reduction in causal inference: analyzing the effect of school achievements on acute complications of type 1 diabetes mellitus. Preprint. *arXiv* 2013; 1309.4054.
- de Luna X, Lundin M. Sensitivity analysis of the unconfoundedness assumption with an application to an evaluation of college choice effects on earnings. *J Appl Stat* 2014;**41**:1767–84.
- Häggström J, de Luna X. Targeted smoothing parameter selection for estimating average causal effects. *Comput Stat* 2014;**29**:1727–48.
- Waernbaum I. Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. *Stat Med* 2012;**31**:1572–81.
- de Luna X., Johansson P. Non-parametric inference for the effect of a treatment on survival times with application in the health and social sciences, *J Stat Plann Inference* 2010; **140**:2122–37.