



The role of quality frameworks to improve the quality of statistics based on administrative registers – The case of organic meat production

Daniel Persson, Ulf Svensson

Swedish Board of Agriculture, Statistics division

Vallgatan 8

Jönköping, Sweden

daniel.persson@jordbruksverket.se, ulf.svensson@jordbruksverket.se

DOI: 10.1481/icasVII.2016.f32d

ABSTRACT

It is an important and difficult task to make sure that the quality of statistics obtained from administrative sources is sufficient for statistical needs. In recent years several studies have been made investigating quality frameworks for administrative registers. These includes the quality of the register itself, the possibilities of integrating administrative registers into statistical registers and how to document the quality of the statistics produced.

The aim of this paper is to discuss how quality frameworks can be used to improve and declare the quality in the statistics obtained from administrative sources. The frameworks will be discussed in relation to where in the statistical production process it can be included. The usefulness of two frameworks will be compared, i.e. the framework published by Statistics Netherlands (Daas et al. 2009) and the framework published by Laitila et al. 2011.

The paper will end in a discussion on the usefulness of the quality frameworks for administrative registers to improve the quality of the statistics and make suggestions for additional aspects that will have to be taken under consideration.

The case of using administrative registers from the control bodies to produce statistics on the structure of organic farms and on the amount of organic production will be used. The case will show how the quality frameworks can be used to discuss the statistical quality when using administrative slaughtering registers to producing animal production statistics. The case will also show how several registers; the registers from the control bodies, the farm register, the cattle database and the slaughter register, can be merged in order to get the statistics required.

Keywords: Administrative registers, Quality frameworks, Organic

1. Background and aim of the paper

It is a difficult decision whether to conduct a survey or derive statistics from administrative registers. This is especially true for official statistics, which must be a source everybody can trust.

A few years ago the demand for statistics about organic farming grew high. The Swedish Board of Agriculture (SBA) faced different alternatives. One alternative was to continue to publish figures from the three different audit bodies that make sure a farmer meets all of the requirements for organic farming. That alternative, however, would mean many drawbacks would persist. The figures were not comparable with the other agricultural statistics that we publish. The figures were shaky when broken down beyond national level. The advantage was that we already had a system to handle the data.

The second alternative was to conduct a sample survey of organic farms. There were, however, some disadvantages to this choice as well. To send out a questionnaire and to register all of the responses would incur an annual cost. We were also considering the burden for the respondents to fill out the questionnaire. The advantage was that the statistics would be comparable with all of the other agricultural statistics.

The third alternative was to derive the statistics from registers. The disadvantage was that it would take us some time to figure out how to do it. Also, we had to make sure that the administrative registers were of good quality. The advantage was that, once done it could be used year after year with minimal costs. The farmers with organic production would not have to fill out an extra questionnaire, hence the response burden would be zero.

After considering the pros and cons, we started to investigate how to proceed with alternative 3, to derive the statistics from administrative registers. We figured that the best way of producing comparable statistics was to link the statistics with the Farm Register (FR). If we could link the Organic Register (OR) with the FR, we could not only derive statistics for the whole country but also for any other geographical subdivision which might be requested. But we needed a framework to put our registers in.

In recent years several studies have discussed quality frameworks for using administrative registers for statistical purposes. The aim of this paper is to discuss the concepts developed by Daas et al. 2009 and Laitila et al. 2011 for the decision whether to use administrative registers to produce statistics about organic production.

2. The case of the statistics on organic animals

In all we found that we had three different registers to use. The first register is the FR, which contains information about all of the holdings in Sweden. Most notably for this exercise the number of animals of different kinds on the first Thursday in June every year. There is also information about social security numbers (SSN) of the holders and of other persons sending information to different registers at the Swedish Board of Agriculture (SBA). There is also information about the Production Location Number (PLN), i.e. the place where animals are held, the address and the phone numbers to the holding. Every third year we conduct a census updating all of the holdings in the register. Among the things that we update are the unique client id for all clients connected to one of the three control bodies certifying organic farming in Sweden. The FR is regulated by Swedish legislation.

The second register is the slaughter register (SR) that contains information about the number of slaughtered animals and their weight. The register contains information about the SSN of the

persons reporting the slaughtering to the slaughterhouses and also from which PLN the slaughtered animals were coming from. The SR is regulated by Swedish legislation.

The third register contains information about holdings with organic production, the Organic Register (OR). This register is a combination of the registers of the three control bodies. The SBA receives these registers and merge them. This register contains information about what kind of organic animals that are raised on the holding. However, the way the animals are counted differs. For some animals it is the average number of animals during the year and for some it is the maximum number of animals during the year. There is information about the SSN, address and phone number of the holder. The holding's client id with their respective control body is also included in the register.

The aim is to combine all of the three abovementioned registers in order to extract the number of organic animals and the amount of organic slaughtering.

2.1 Setting up the frame

The idea for the number of organic animals is that we would match the OR with the FR. If a holding had organic cows in the OR, then all of the cows in the FR would become organic. And if a holding had organic pigs in the OR, then all of the pigs in the FR would become organic. We had to make one exception for poultry, because the different kinds of poultry are so different. If a holding had organic broilers in the OR then the broilers in the FR would become organic, but all of the other types of chickens would not become organic.

One of the big advantages with using the number of animals in the FR is that the data is fully comparable with the other figures from the FR. With the FR we can also determine an exact geographical point where the animals are located.

Some of the difficulties were that one holding in the FR could be two holdings in the OR, so there was not a one-to-one ratio. Some manual work was required to establish a link between the holdings in the OR and the corresponding holding in the FR, as this could not be done automatically. Overall the matching was not much of a problem though.

The idea for the organic slaughtering was the same as the idea for the organic animals, but we also had to include the SR. If a holding had cows that were present in the OR and also had slaughtering of cows, then all of the cows slaughtered would become organic.

Both Laitila et al. and Daas et al. contains suggestions of a system of indicators for quality assessment of administrative data, so that one can get an overview of the registers. On indicator level the concepts of Laitila et al. and Daas et al. are to a large extent overlapping. However, there are some differences in the way indicators are systemized and how quality is assessed. Daas et al. sees three dimensions: the administrative register in itself, metadata about the source and the data in the source. Indicators are developed within each dimension. Laitila et al. discusses indicators in three dimensions: output data quality, input data quality and production process quality. However, they group the indicators related to the work process, i.e. indicators related to information from the administrative authority, indicators related to data editing of the source, indicators related to integrating the source with the statistical register and integrating the survey with similar variables.

The indicators in the hyper-dimensional described in the text are summarized in table 1. The SBA is the supplier of the data sources of the FR and the SR. The supplier of the OR are the three control bodies that supplies one separate register each to the SBA.

Table 1: Evaluation of registers according to the hyper-dimensions of Daas et al. 2009 and Laitila et al. 2011

	Farm Register	Slaughter register	Organic register
Evaluation of registers according to Daas et al. 2009			
Source dimensions			
S1. Supplier	+++++	+++++	++++
S2. Relevance	+++++	++++	+++
S3. Privacy & security	+++++	+++++	+++
S4. Delivery	+++++	+++++	+
S5. Procedures	+++++	++++	++++
Metadata dimensions			
M1. Clarity	+++++	+++++	++
M2. Comparability	+++++	+++++	+
M3. Unique keys	+++	++++	+++
M4. Data treatment	+++++	+++++	+
Information from the administrative authority according to Laitila et al. 2011			
Relevance			
A1. Relevance of the population	+++++	++++	+++
A2. Relevance of units	+++++	+++	+
A3. Relevant matching keys	+++++	++++	+++
A4. Relevance of variables	+++++	+++	++
A5. Relevance of matching time	+++++	++++	++
A6. Study domains	+++++	+++	++
A7. Comprehensiveness	+++++	++++	++++
A8. Updates, delivery, punctuality	+++++	+++++	++
A9. Comparability over time	++++	++++	+++
Accuracy			
B1. Primary keys	++++	+++	++
B2. Quality of reference variables	++++	++++	++
B3. Doublets	++++	++++	++
B4. Missing values	+++++	+++++	++
B5. Wrong values	++++	+++++	++
B6. Output of preliminary data	+++	++++	+

What we found is that the FR and the SR are of very high quality. The OR, however, does not have the same quality standards. The FR is the base in Swedish agricultural statistics and contains all of the base variables. It is updated yearly.

Regarding the dimension of supplier (S1) the FR and the SR are situated at the SBA. The OR is sent to the SBA from the three control bodies. The knowledge of the OR is OK. Regarding the dimension of relevance (S2), the relevance of the units (A2) and the relevance of the variables (A4), the slaughter statistics is a register where the slaughterhouses report the number of animals that have been slaughtered and their corresponding weight each day. However, the register does not have any information of the animals slaughtered on the farms. The OR stems from the audit bodies that control that a farm meets the requirements of certified organic production. However, the OR contain a few variables of interest to us, namely what kind of animals are certified as organic and information to conduct our matches with other registers.

The privacy and security (S3) for the FR is very good. Nobody outside the statistical department at the SBA has access to the data. It is also protected by Swedish law. As for the SR nobody outside the SBA has access to the data. The situation is somewhat similar for the OR. Addresses might be disclosed but the number of animals will not be disclosed.

Regarding the delivery (S4), procedures (S5) and updates & punctuality (A8), then for the FR and the SR there is not much of a problem, because these registers are situated at the SBA and historically they have always been delivering data in time. There are usually no major errors in the material and if found, they are corrected swiftly. We are of course very familiar with how the data is collected and if we should have any questions they will be answered in a short period of time. The OR, however, usually requires more work. When the registers arrive they must be checked for errors and requests must be sent to the control bodies for them to correct the errors.

For the clarity and comparability (M1, M2) and the relevance of the variables and time (A4, A5), the base variable is holdings in the FR. Linked to these holdings are the number of animals from the FR and the number of slaughtered animals from the SR. The auxiliary variables come from the OR, i.e. what kind of organic animals are on the farm. In the FR the number of animals are counted on the first Thursday in June, but in the OR the number of animals could be either the maximum or the average during the year. Regarding the relevance of time, there should not be any problem, because all of the registers have strict codes regarding when a year begins and ends.

As for the unique keys (M3), primary keys (A1) and doublets (B3), there is some concern for all of the registers, because we know there are doublets in the material for the FR and the OR. For the FR, this is less of a problem in the census years and more so in the years between censuses. Holdings with organic production can choose between one of the three accredited control bodies and sometimes they change control body between the years. Since the registers of the control bodies are not fully up to date, one holding can be present in more than one register. As mentioned before, the holding is the primary unit in the FR, but in the other two registers, holding is not a primary key but instead a person. In some cases, two or more units from the SR or the OR must be put into one holding in the FR.

Data treatments (M4) are conducted on the FR every year before it is being published. There are checks whether the holding is part of the population, checks for extreme values, do all holdings have a holder, and is the data consistent and so on. For the SR, the holdings should report the animals that are being slaughtered. The same information should also be reported by the slaughterhouses. This form of double reporting should indicate that there are very few posts that are faulty. The OR is kind of different, because the audit bodies do not seem to put that much effort into data treatment. There are not many missing values (B4) but there could be substantial amounts of wrong values (B5). This of course, is a burden put on the SBA.

The third hyper-dimension introduced by Daas et al. 2009 is the data dimension that focuses on the quality aspect of the data in the source. Laitila et al. 2011 focus on the indicators of input

data and production process quality. Hence Laitila et al. 2011 focus on register level, but Daas et al. 2009 assume that we base our analysis on the overall performance combining all of our registers and in the end receive the statistics. The quality is therefore assessed on the whole approach and not only on the registers themselves (Table 2).

Table 2: *Quality of the data for statistical purposes according to Daas et al. 2009 and Laitila et al. 2011*

Daas et al.	Model of the number of organic animals and organic meat production	Laitila et al.	Model of the number of organic animals and organic meat production
Q1. Technical checks	+++	D1. Relevance of variables as such	++++
Q2. Over coverage	++++	D2. Relevance of variables as auxiliary information	++++
Q3. Under coverage	++++	D3. Relevance of variables to improve statistics	++++
Q4. Link ability	++++	D4. Under coverage in administrative registers and in FR	++++
Q5. Unit non-response	++++	D5. Over coverage in administrative registers and in FR	++++
Q6. Item non-response	+++	D6. Duplicates in administrative registers	++++
Q7. Measurement	++++	D7. Wrong units in administrative registers and in FR	+++
Q8. Processing	++++	D8. Missing values	++++
Q9. Precision	++++	D9. Wrong values or wrong units, cross section data	++++
Q10. Sensitivity	++++	D10. Wrong values or wrong units, longitudinal data	++++

Regarding the technical checks, there are no problems reading the data of the FR and the SR. The OR comes from the three audit bodies and is of worse quality. However, the data is fully readable after some extensive manual labor (Q1).

The variables in the OR are used as auxiliary variables and variables to improve the statistics (D2, D3). From the register we know which holdings are conducting organic farming. Since a holding can be both organic and conventional, we also use the information of what animals are organic. So if a holding is raising organic cattle, then all the cattle on the holding and all of the slaughtered cattle are assumed to be organic. In our model, it has worked out very well and the information has been very useful.

Over coverage in the SR is estimated to be negligible because of the double reporting conducted by the farmer and the slaughter houses. In the OR there could be over coverage if a holder is not removed from the register even though they have ceased producing organically. However, the over coverage is assumed to be very small. Over coverage in the FR is negligible for

the years when we conduct a census (Q2, D5). For the years between there can be some over coverage since we do not send out questionnaires to every holding. It must, however, be emphasized that the over coverage is small and will not lead to any big overestimation. For all three registers combined, the over coverage is assumed to be very small.

The under coverage is assumed to be small for all of the registers and the linkability is assumed to be high (Q3, Q4, D4). The newly started holdings almost certainly will apply for subsidies and will therefore be incorporated in the FR. Less than 0.1 % of the animals in the SR are not linked to the FR now. For sheep there has been a few years where we could not link as much as 1 % of the slaughtered animals. In the OR there are usually 1-3 holdings that we cannot link, but those holdings could very well be closed down. There must be some mismatches between the registers, but we do not believe that the number is very high. It is of course hard to give a specific number, but from what we have seen, it is most certainly quite low.

Regarding the unit non-response and the item non-response (Q5, Q6, D8) we have a response rate of about 97 % in the FSS and in the livestock surveys. In the SR the response rate should be very close to 100 % because of the double reporting and in the OR the response rate should also be very close to 100 % because otherwise the farmers will not receive any subsidies or be able to sell the meat as organic. The item non-response, however, is harder to measure since we do not know if a holding has forgotten to fill in the number of sows under the pig section in the FSS or the livestock survey. For the bigger holdings we might find out that information is missing, but for the smaller ones it is very difficult.

The number of organic animals in our model is compared with the number of organic animals reported by the Swedish organization KRAV. The number of organic animals are about the same in both our model and KRAV:s figures (Q7). We also compare the amount of organic meat with other organizations to see if our figures are in line with them. There can of course be some small random errors. The systematic errors are assumed to be small, but we are underestimating the number of slaughtered animals due to the slaughtering on the farms. However, the slaughtering at the farms are close to insignificant.

In the processing (Q8) there are almost no adjustments and imputations made, because we see the FR and the SR as already finalized. Hence, there are no checks for outliers. In the OR there may be occasional adjustments and imputations made. Since we have figures for all of the holdings, there is no standard error in our estimates (Q9). However, when we made a time series from 2009-2013 for the variables involved, we noticed smooth trend lines (D10).

In the process we have made comparisons between the OR and the FR (D1, D9). The data in the two different registers are not totally comparable since the FR measures the number of animals the first Thursday in June every year and the OR measures animals differently. For example, the OR measures the total number of broilers produced during the year. However, in this case we divided the number of broilers with the production time and received a number that we could compare with the FR. Our comparisons turned out well and even though there could be differences, we concluded that the material from the OR was of good quality.

In our model there could be missing values or duplicates in the OR, but we believe that the numbers are small (Q10, D6). The total bias of the totals is assumed to be low.

2.2 *What is the result?*

After some testing and evaluation, we came to the conclusion that we had a very good model. The figures are in line with other sources and we can easily update the model each year with new figures. The response burden on the farmers is zero, because we use administrative registers and the cost for us is very low since we do not need to make another survey.

3. Discussion

The two frameworks have proven to be an excellent help in deciding the direction of the work with organic meat production. We wanted to use registers to obtain the organic meat production, but we weren't sure of if the registers were good enough. Both of the frameworks are thorough and takes every detail into account. They are also an objective and a standardized way of determining the quality of the registers that you are about to use. They have helped us to really consider what options there are, as well as declaring the quality of statistics obtained from administrative sources.

We found that the FR and the SR were registers of good quality and with the two models by Daas et al. 2009 and Laitila et al. 2011 we could quantify just how good they really are. However, these registers are maintained by the SBA and are thus in our own control. The OR was less good, but sufficient enough to be used in our model. Since the OR is not in our own control, we depend on the three organizations that maintain the register, and as we could see, there were some problems with the register. When we combined all of the registers however, the framework proved to us that our model was of very good quality. We were also able to see where the model could be improved.

The two frameworks are similar. One advantage of the framework of Daas et al. 2009 is the focus on the procedures for accessing data and the legal prerequisites. One advantage of Laitila et al. 2011 is the focus on the production process indicators as a separate step.

It is obvious that the quality frameworks can be used to determine whether to use administrative registers or not. For us the frameworks have proven to be very helpful in our decision-making.

REFERENCES

- Daas P., Ossen S., Vis-Visschers R., Arends-Tóth J. (2009). Checklist for the Quality evaluation of Administrative Data Sources. Statistics Netherlands
- Laitila T., Wallgren A., Wallgren B. (2011) Quality Assessment of Administrative Data, Research and Development – Methodology reports from Statistics Sweden 2011:2
- Wallgren A., Wallgren B. (2014). Register-based Statistics: Statistical Methods for Administrative Data. Second edition, Wiley