Journal of Official Statistics
Vol. 8. No. 4, 1992, pp. 481–498

Statistical Defensibility as Used by U.S. Department of Agriculture, National Agricultural Statistics Service

Rich Allen[1]

Abstract: This article outlines steps that have been taken to ensure greater defensibility of statistical estimates and products by one organization. Illustrations are given of changes in procedures that have been made along with examples of procedures that were examined and not changed. Specific references are made to data confidentiality and publication concerns.

Key words: Statistical standards; data confidentiality; survey administration.

1. Introduction

This article outlines the origin of the concept of statistical defensibility within the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture. A number of examples of specific applications of the concept are described. In addition, the relationships of statistical defensibility and quality management are developed.

For most people, the first use of the term "statistically defensible" came in a ruling by Judge Horace W. Gilmore in the case *Young v. Klutznick* when he ordered the United States Census Bureau to adjust the 1980 U.S. population census counts (Gilmore 1980). He stated that methods of adjustment shall remain within the discretion of the Census Bureau "as long as they are statistically defensible." However, Judge Gilmore did not define statistical defensibility, nor was there a standard definition available in the literature.

Several notes and opinions on statistical defensibility were published in The American Statistician in August 1982. As part of his comment, Wolter (1982) stated that the Census Bureau, in its written reply to Judge Gilmore, defined seven criteria that should be used in addressing the statistical defensibility of a statistical procedure:

1. appropriateness to the use to be made of the results
2. provision of measures of uncertainty
3. listing and verification of assumptions; demonstration of robustness
4. description of data sources, reliability, and limitations
5. reproducibility
6. timeliness
7. least cost for given degree of uncertainty.

While NASS has not defined its own list

[1] Deputy Administrator for Programs, National Agricultural Statistics Service, U.S. Department of Agriculture, Room 4133 South Building, Washington, D.C. 20250–2000, U.S.A.

The views expressed in this article are those of the author and do not necessarily represent the policies of the National Agricultural Statistics Service (NASS).

of criteria or formally adopted the Census Bureau list, this article should show that the agency's major criteria in its utilization are numbers 1 to 6 in the census list. (Since government statistical agency budgets are quite tight, the cost factors of number 7 are also always a concern.) The demonstration of robustness in criteria 3 is particularly important to NASS as is criteria 1 which emphasizes a knowledge of the audience which will use the information.

The following definition of statistical defensibility summarizes the NASS application of the term:

> "Using the most appropriate survey, analysis, and estimation procedures which are practical for each statistical application."

Statistical defensibility is thus broader than statistical methodology. While choosing the most appropriate methodology from among those available for any particular application, the NASS concept will also encompass data dissemination, survey design, questionnaire design, and improvements in editing and analysis procedures. NASS has developed a wide range of standards for nearly all of its activities as part of its statistical defensibility efforts.

2. Origin in NASS

The term "statistically defensible" (in its various forms as a noun or adjective) has been used by NASS since 1983. Its origin was in deliberations among the authors of a long-range plan for the agency which was published as *Framework for the Future*, although the term is not used in the publication itself (Statistical Reporting Service 1983a). *Framework* called for the agency to develop and operate within a set of standards for every aspect of the agency's operation. The authors were particularly interested in the development of procedures for the interpretation of combinations of probability and *nonprobability* based data which would be "repeatable" and "defensible."

One of the earliest actions resulting from implementation of the recommendations of *Framework* was the formation of a task force to review the estimating procedures of the agency. That task force's report entitled *Crop Reporting Board Standards* in July 1985 did use the term statistical defensibility (Statistical Reporting Service 1985). The report states: "SRS (the agency acronym at the time) ... should also publish estimates that are statistically defensible."

There were two major concerns about statistical defensibility during the deliberations which led to the *Framework* recommendations. The first was that, although the agency had implemented a number of probability surveys following its last long-range plan in 1957, many major statistical series still depended largely on nonprobability indications. One point of view expressed was to eliminate nonprobability surveys although that would imply that large increases in funding would be needed or the program of estimates would need to be reduced.

The second concern was that the agency employed a review board approach which used expert judgment for nearly all major reports, even those based on probability conducted surveys. While this approach to "setting" estimates had long been a hallmark of the agency, one point of view discussed was to recommend that estimates be set by a statistical calculation if probability indications were available.

These aspects and the resulting decisions are discussed below.

3. Statistical Defensibility and the Agricultural Statistics Board

The U.S. Department of Agriculture (USDA) since 1905 has utilized a review board approach for major reports in which

an independent review of state level indications and recommendations is performed (Statistical Reporting Service 1969). Up to 1905, estimates were finalized and published by a small number of individuals located in Washington, D.C. Judgement was used to interpret the reports and observations forwarded from the individual states. Concentrating the final authority in such a small number of individuals plus relatively lax security procedures presented a great potential for abuse. It was found in June 1905 that one agency employee had been leaking advance information to an outside accomplice. When rumors of a leak tightened some precautions such as locking the rooms where estimates were being finalized, the Associate Statistician involved worked out a set of signals which involved raising or lowering the window shade.

The scheme was only discovered when the outside accomplice charged that the June 1905 cotton report must have been falsified since the direction of the report did not match the signal. (In fact, the inside person had not been able to send a signal of a late change which occurred on that report.) Besides the removal from office of the individual who was doing the signalling, new physical security procedures (including sealed blinds) were implemented. In addition, a "board" review procedure was implemented by the new head of the estimating procedures. The board later was specified in legislation which requires that individuals from State Statistical Offices must participate in the final deliberations on the monthly *Crop Production* report. Legislation went as far as stating that at least three field representatives from cotton producing states must participate in all cotton production forecasts. The laws governing crop reports are summarized in an agency publication entitled *Scope and Methods* (Statistical Reporting Service 1983b).

The board approach ensures an unbiased review and provides input from knowledgeable staff members as statisticians from State Statistical Offices serve, on a rotating basis, on the board. The board review was originally referred to as the Crop Reporting Board (now the Agricultural Statistics Board). Despite the original name of Crop Reporting Board, the board review process was also utilized for other reports such as hogs and cattle.

One ongoing justification for the use of a board review process by NASS is that nearly complete utilization data become available for many commodities at some time during the next production year. Good quality data are available on exports of both crops and livestock and on livestock slaughter. Also, crops utilizations such as soybean crushings and wheat millings are reported by the United States Bureau of the Census. These disappearance data are valuable for revisions of crop production and livestock inventory estimates but they are also helpful on a current basis for setting periodic estimates of grain stocks on hand.

The approach that NASS uses is the calculation of a balance sheet which estimates total supply, then subtracts out disappearance data to obtain a second indication of present inventory to match against survey indications. Table 1 illustrates a December 1 to March 1 balance sheet for hog inventory. The commercial slaughter, imports, and exports information are from administrative sources. Pig crop, deaths, and home slaughter come from NASS surveys.

If a balance sheet is far out of balance, it implies that one or more of the factors of beginning inventory, pig crop, or ending inventory are at an incorrect level. Slaughter during the past period and the pig crop which provides information on the amount

Table 1. U.S. hog balance sheet: December 1–March 1

Item	1989	1990	1991	1992
On hand December 1	55,469	53,821	54,477	57,684
Dec–Feb pig crop	21,168	20,362	21,325	23,183
Imports	319	298	225	205
Total supply	76,956	74,481	76,027	81,072
Commercial slaughter	22,069	21,662	21,643	23,599
Farm slaughter	145	140	120	100
Deaths	1,250	1,100	1,160	1,100
Exports	91	31	48	45
Total disposition	23,555	22,933	22,971	24,844
Indicated March 1	53,401	51,548	53,056	56,228
Estimated March 1	52,965	51,150	52,760	56,110
Difference	−436	−398	−296	−118

of breeding stock the previous year give insight on whether the previous inventory was correct. NASS does *not* adjust figures to completely eliminate the imbalance or residual. Every item in the balance sheet is subject to nonsampling errors and the inventories, pig crops, deaths, and home slaughter are affected by sampling errors. For those reasons, NASS prefers to allow a reasonable imbalance rather than "forcing" a zero residual.

Instead of trying to replace the Agricultural Statistics Board procedure by weighting and analysis of probability survey indications, the writers of *Framework* recommended improvement of the review procedures of the board. Many of the indications that NASS receives from sample surveys, particularly early season yield forecasts, are biased so the agency needed to develop more consistent procedures for interpreting the biases.

One major improvement in procedures implemented since the *Framework* and *CRB Standards* reports has been to better formalize the interpretation process. Figure 1 illustrates a not atypical relationship between Farm Report (a farmer survey) yield indications, Objective Yield (a probability selected in-field survey used in the states with significant planted areas of the crop) indications, and actual end-of-season yield levels. This example is based on data for the region which consists of all major producing states so year-to-year consistency is greater than for individual states.

The historic agency review procedure was for each state office to conduct their Farm Report and Objective Yield surveys and to review time series charts of past performance in making their recommendations. Those recommendations from all states were then reviewed by the Agricultural Statistics Board. For the most important or "speculative" states, which receive extra security procedures, this review started with review of regional levels before reviewing state data. If the statisticians in a particular state were particularly strong supporters of the probability survey approach, their recommendations might tend to be fairly close to the objective yield indication. Conversely, if the statisticians favored the more traditional farmer survey approach with more samples, they might recommend closer to the Farm Report indication upon reviewing the same data. Table 2, which lists the data which are graphed in Figure 1, indicates

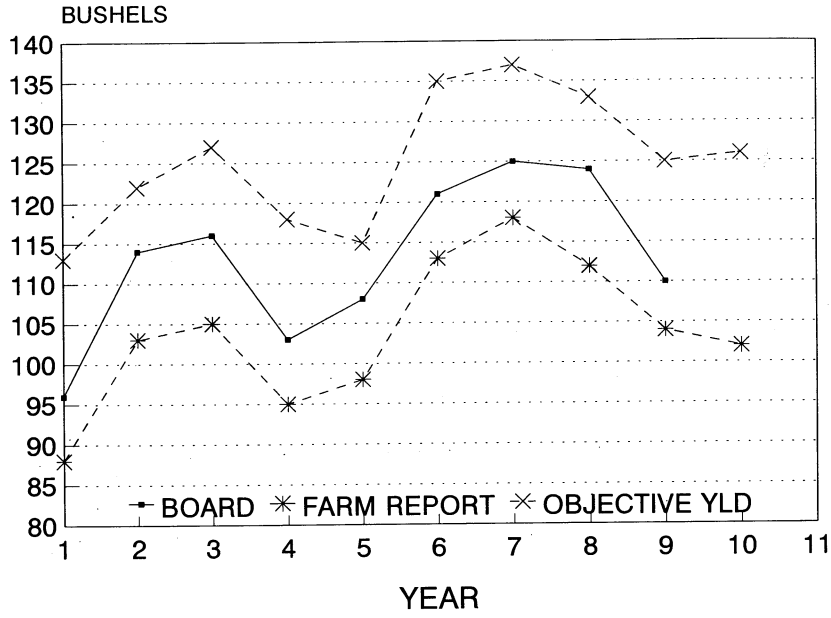

Fig. 1. Corn yield indications and final board level

that the survey indications for the current year (year 10) differ by 24 bushels per acre. This difference lends itself to much inconsistency in interpretation.

The approach implemented is referred to as the Comparison Table. In Table 2, the year-to-year departures between indications and final actual yield are calculated and averaged over the previous nine years. These average differences for each indication are then added to the current survey indications to form an adjusted indication for each data source. Notice in Table 2 that the difference between the two adjusted figures is now only

Table 2. Comparison table: Corn yield indications and differences from final board

Year	Board	Farm report	Difference	Objective yield	Difference
1	96	88	8	113	−17
2	114	103	11	122	−8
3	116	105	11	127	−11
4	103	95	8	118	−15
5	108	98	10	115	−7
6	121	113	8	135	−14
7	125	118	7	137	−12
8	124	112	12	133	−9
9	110	104	6	125	−15
10	?	102		126	
Average difference (years 1–9)			9		−12
Computed yield		111		114	

three bushels. In this case, with the objective yield indication showing a slight increase over the previous year and the Farm Report being lower, a season with poor early season weather might have occurred. The objective yield is indicating a high plant count and good yield potential if fruit weight is close to normal but farmers are pessimistic based on the less than favorable weather.

Board members can now focus on other factors which determine their own recommendation within this indicated spread of three bushels. Since this process is used for monthly yield forecasts during the growing season which start about four months before harvest, the present recommendation may be tempered by the previous month's forecast level. Although the agency forecasts yields based on first of the month indications, weather conditions since the date that the surveys were taken may receive some consideration in final deliberations. Members may also examine years in the previous nine which had similar weather conditions to the present year and give those years more weight in their recommendation.

Granted, the Agricultural Statistics Board could mathematically combine adjusted indications based on the performance of the last nine years and adopt this calculated composite. However, there is the opportunity to use expert judgment about the current season by presenting the full set of data and the indicated adjustment.

Using this approach in both the state offices and the Agricultural Statistics Board has resulted in more consistent interpretations and a reduction in the number of times a state recommendation is changed by the top down board review process.

Table 3 lists the number of changes to state recommendations for 1982 (before the comparison table approach was introduced) and 1987 (when the approach was in place but analysts had not looked at any year-to-year differences). The table includes four months of corn, soybeans, and wheat forecasts and five months for cotton.

Table 3. Frequency of state yield recommendation changes[1] 1982 versus 1987

	1982	1987
Corn	14 of 28	7 of 28
Soybeans	16 of 32	6 of 32
Wheat	8 of 32	7 of 32
Cotton	15 of 30	7 of 30

[1] Counts based on major states for the most critical months of the season.

The value of using expert judgment in the ASB process was particularly shown in 1988 when a very severe early season drought was encountered. Other droughts or widespread disease problems had tended to occur after August 1 but the 1988 early season conditions were very seriously affected. The farmer locality yield and objective yield corn indications for August had differed by an average of 21.1 bushels the nine years previous to 1988 with a minimum difference of 12.2 bushels and a maximum of 30.0. The 1988 indications varied by 55.5 bushels for the two surveys. Applying the usual bias adjustments to the two indications still resulted in a difference of 34.4 bushels in the two adjusted indications.

While board members knew that there is a human tendency to always overestimate the affects of a disaster such as a flood or a drought, they also knew that the 1988 objective yield indications, based on historic ear weights, were too high. After deliberation, and review of individual objective yield model components, the board adopted average yield for the speculative states in August 1988 was 80.0 bushels per acre compared to the 94.6 bushels per acre that would have been generated by giving each historic relationship equal weight. That figure and the resulting U.S. level average yield of 78.5

were initially criticized as being too high. The level proved to be quite reasonable, however. While the drought was never "broken" during 1988, conditions stopped deteriorating and no additional adverse conditions were experienced at harvest with resultant final average yields per acre of 85.8 in the speculative states and 84.6 for the U.S.

4. Statistical Defensibility and Nonprobability Surveys

With the vast number of reports of the agency (some 400 reports are issued each year including quarterly, monthly, and weekly series as well as annual publications) and the quick turn around from start of data collection to publication, well managed nonprobability surveys have played an important role. It should also be pointed out that nearly all agency surveys are collected on a voluntary basis; thus, it has been important to cultivate good communications with respondents.

Some of the surveys usually included in the nonprobability category have actually been total, or nearly complete, enumerations. Many of the crop commodities estimated by NASS are marketed through only a few channels in a state, or are produced under contract to a few firms, or producers receive seed from only a few outlets. Thus, the entire crop can be accounted for, with a fairly high certainty level, with only a few contacts. This historically has not been regarded as a probability survey process because every unit does not have to be accounted for in each survey.

A more common nonprobability survey application for NASS was the monthly Farm Report. This was a mailing list of farmers who received a questionnaire each month with the appropriate questions for that time of the season. This panel survey usually included at least one reporter from each township in the state. Individuals stayed on the mailing list as long as they continued to report on a reasonably frequent basis. If they stopped reporting or left farming they were replaced by someone from the same area of the state, probably someone who had been a good reporter to other similar, but not as frequent, mail surveys.

Many "farm reporters" took great pride in providing this service to agriculture and might go for several years without missing a report. This high response rate and the relatively stable makeup of the panel allowed agency statisticians to monitor performance of current reports against final end-of-year yield levels.

The agency actually made many changes in the Farm Report survey over time. (One might conclude that these changes were made in the name of better statistical defensibility but were not documented as such at the time.) When the Farm Report was begun, farms within states in the United States tended to be very similar. They depended mainly on animal power so every farm grew some feed for those animals. Farms were quite self-sufficient for their own food so most farms had some vegetables, poultry, and dairy. Cropping practices tended to be very similar from farm to farm. The monthly Farm Report was used as the major survey vehicle for dairy, poultry, and farm labor estimates in addition to crops data. The changing monthly questions also collected data on items from firewood to land values to livestock values.

As farms became more mechanized, they tended to become more specialized. Farmers concentrated on producing only a few crops and many farms stopped producing dairy and poultry. As these changes occurred, numbers of positive reports dropped for many items and indications were not as

consistent. The agency then shifted to single-purpose surveys for specific livestock items and farm labor. Some other items were also deleted, such as fruit, until the Farm Report became primarily a field crops only survey.

While the Farm Report continued to provide serviceable yield indications, there were concerns that the reporter self-selection procedure could not be defended. There also was no way of calculating the statistical precision of the indications.

Several factors had to be considered in testing an alternative to the Farm Report for monthly yield indications. In many cases, during the first month of the forecasting season, questions on condition of the crop were asked instead of yield since it had been believed that individuals would be able to better report condition than to interpret a yield. Also, the reporters were usually asked to report a locality yield rather than their farm yield, both to increase the total number of reports and to counter a feeling that reporters might not be willing to "stick their neck out" in giving early season interpretations of their fields.

Efforts to develop an integrated quarterly survey program which combined already existing probability hogs and pigs and grain stocks surveys with appropriate crop acreage and end-of-season production data started in three states – Illinois, Tennessee, and Arizona. Once the quarterly surveys were underway, samples of operations which reported the crops of interest were selected for monthly yield questionnaires.

Experience in these states showed that farmers would report yield expectations for their own fields early in the season. The results also showed that early season underreporting biases were essentially the same as experienced in the Farm Report. Thus, these states switched over to the new procedure and, by 1992, all states have been converted. Sampling for small grain crops comes from acreage intention reports in March; other field crops come from the June Agricultural Survey which measures mostly actual plantings. The first monthly yield survey rechecks the planted area and area intended for harvest as grain. Weighted average yields are calculated based on probability of selection of each operation. The same operations are surveyed each month of the forecast season.

5. Progress to 1988

Because no agency instructions or guidelines were issued on statistical defensibility, debate and discussion continued on the meaning within NASS. One topic at the April 12, 1988, meeting of the NASS Program Planning Committee was a presentation on statistical defensibility.

The approach in that presentation was to define that:

"Statistical Defensibility is *Not* Just:
- Publishing Sampling Errors
- Calculating Additional Estimators
- Designing Nicer Questionnaires"

"Statistical Defensibility Involves:
- Planning
- Analysis of Data
- Documentation
- Review of Possible Improvements"

Also included in that presentation was a listing of activities that the agency had implemented or was in the process of researching that would lead to greater statistical defensibility:

1. Shifting to collection of inventory data on a post reference date basis. Data for first of a calendar quarter reports on grain stocks and hog inventories were formerly collected mainly before the first of the month with no instructions whether reports were to

measure inventory as of that date or expected as of the first of the month.

2. Standardizing wording for all probability prices received questionnaires with documentation of valid exceptions. While questionnaires were all derived from a master questionnaire, individual State Statistical Offices versions varied in how much instruction was provided on definitions of "includes" and "excludes" and other factors which could affect results.
3. Defining which crops must be covered in crop progress tables and standardizing crop condition adjectives. The Weekly Weather Crop Survey, a relatively small nonprobability survey which results in a publication the first working day of each week from planting through harvest, measures crop conditions as well as crop progress. Until about 1986, State Statistical Offices controlled when they started and stopped asking condition questions on each crop. Also, various four- or five-level systems of adjectives were used. The new approach defines standard adjectives (excellent, good, fair, poor, and very poor) and specifies when states must begin asking the questions. As a result, standard tables are summarized each week giving state level and major state weighted condition data.
4. Sampling operations not represented by the current list frame for crop acreage and production surveys in order to form true probability indications. Formerly, only one full probability indication was available each year – the acreage indication from a nationwide area sampling frame survey conducted in June. Probability indications are now available for each of four quarterly surveys during the year.
5. Using administrative data and past survey performance information to interpret current survey indications and set estimates. "Balance sheets," utilizing administrative data such as slaughter, exports, seed used for planting, wheat millings, etc., had long been used for annual revisions. The new emphasis is on expanded use for surveys throughout the estimating year.
6. Using "presence/absence" information for nonrespondents to improve survey indications. Someone refusing to answer the questions on hog production is likely to actually have some hogs; if not, it would be easier to just report zero than to refuse. The current approach is to determine, if possible, whether each refusal or inaccessible has hog production. If they do, they are imputed for by the average of all positive reports in that stratum, not all reports including zeroes.
7. Publishing past performance comparisons (such as root mean squared errors) as guidelines of expected performance for data users. NASS conducts a variety of different statistical surveys, including many that forecast across several months of the growing season. The best measure of possible performance in the current forecast season may be how well a series has performed in the past.
8. Developing analysis packages to provide a better review of current survey data relationships. As more probability surveys were developed, interpretation was improved by calculating the effect of expanded data on final indications and by highlighting records which demonstrate unusual relationships such as grain stocks increasing

instead of decreasing during the marketing year.

9. Developing automated outlier detection procedures for use in calculating objective yield model parameters. All objective yield models are based on analyses of historic data of the same maturity code and month compared to end-of-season production levels. Some month–maturity stage combinations have few reports and an undetected outlier can have a great influence in parameters for the current season.
10. Expanding nonsampling errors research such as studies of the effects of using previous survey data and of different recall periods. Many of the NASS survey programs involve multiple contacts during the estimating year. For the most part, previously collected data are not used in the current survey because of concerns about biasing results. However, use of previous data can have a positive effect on respondent cooperation and might be helpful in minimizing biases from other informed respondents when the farm operator is not available.

Reexamination indicates that at least three more activities which were underway could have been added to the 1988 list. They are:

11. Initiating formalized technical reviews of the operations of State Statistical Offices. Since NASS has a national program implemented at the state level and a functional form of organization, review of State Statistical Office procedures for major programs provides information on headquarters' instructions as well as local operations. Some core operations, such as list frame procedures, are studied in each review but commodity level topics vary from cycle to cycle. A procedure was built in to ensure follow-up responses and action on each weakness discovered.
12. Developing a series of Policy and Standards Memoranda which documents the procedures for all aspects of the agency's operations. Policies governing the operations of the agency were formerly found in at least three different series and there was no standardized procedures for review or updating. There is now one series which quantifies policy and standards for nearly all aspects of agency operations and which sets an every five-year review requirement if no updating has occurred.
13. Writing and maintaining an Estimation Manual which standardizes and documents instructions for all commodities and reports. Instructions for various estimates had been found in a variety of locations. Many had not been updated as survey procedures had changed. They also contained considerable redundancy. The new Estimation Manual provides an improved set of instructions for State Statistical Offices.

6. Matching Defensibility to the Audience

One of the points discussed by the authors of the *Framework* was whether the agency should publish sampling variation estimates for all probability based data. There was also considerable discussion of publishing survey indications along with the adopted forecasts or estimates. Another means of providing more information might be to publish ranges rather than point estimates.

Work on the *Framework* report included

interviews of a wide variety of NASS data users. These included analysts and economists with major agribusiness companies, farm organizations, and commodity organizations; the farm media; university and government economists; transportation companies; and individual producers. Identical questions were asked of all individuals contacted, including present users of data, suggestions for improvements, and their evaluation of different presentations of estimates.

One individual provided a good summary of the opinions of the vast majority of people contacted. His statement was essentially:

> Data users consider your numbers as being the best available. If you start publishing sampling errors, even if you improve your survey procedures, most users will probably conclude you now have "errors" in the estimates where you didn't before.

The general feeling from data users concerning the publication of ranges might be summarized as:

> Give your best interpretation. The whole agricultural community uses your numbers and they all have the same number. If you publish a range most will use the midpoint but others will use the lower or higher end which will cause confusion.

To put the views of most NASS data users in other terms, we may say that "users are not interested in precision. They are interested in performance." NASS puts out the same reports on the same schedule year after year. Forecasts of crop production are made monthly starting about four months before harvest. Data users want to know how much confidence to place in a current forecast compared to the previous forecast and to the next one.

These views from data users, instead of being nonstatistical, are actually very practical. Early-season crop forecasts are subject to greater forecast errors than sampling errors. Crop conditions after data collection may degrade due to drought, high temperatures during pollination, or excess moisture and high winds. In other years, conditions may actually be ideal and much better than "normal" which will increase yield potential in subsequent months. No weather models have been found which are consistently helpful in predicting actual conditions and improving forecasts.

Another source of opinions concerning the publication issues came from three independent, but simultaneous, study groups which spent two days with agency officials in 1984. One group included representatives from the American Statistical Association (ASA), one included American Agricultural Economics Association (AAEA) representatives, and the third was composed of three statisticians who had done a more thorough agency review five years before.

All three of the groups shared the opinion of publishing information which allowed the evaluation of the current statistics but the specific recommendations varied. The AAEA group recommended detailed descriptions (Chern, Hushak, Jordan, and Bullock 1984).

> "We recommend that SRS publish all information it has regarding the statistical properties of its estimates. In addition, a detailed description of the processes and procedures of estimation should be published. Information about statistical properties and procedures should be regularly updated. We recommend that SRS consider releasing its survey results as a point estimate with appropriate confidence ranges specified."

The ASA group emphasized the publishing of error approximations, perhaps

through some innovative presentations (Hildreth, Finkner, and Goldberg 1984).

> "Every estimate or forecast should appear with a responsible approximation to error. In some cases this can be an error based on sampling theory, in others both statistical and possible specification errors must be considered. When estimates or forecasts contain important personal judgments, a good approximation to error may involve looking at historical errors for estimates or forecasts made by similar procedures. If the procedures have recently been revised, the approximate error may itself be based partly on personal judgments."

> "It was suggested that including errors would confuse users and erode confidence in data supplied by SRS. If so, this may mean some users have undue faith in the data's accuracy. SRS should study the effectiveness of alternative presentations and explanations. Note that the combination of a point estimate and an interval estimate allows for some expression of suspected skewness in the distribution of the error (the point estimate need not be the center of the interval estimate) whereas a point estimate plus an approximate standard deviation does not."

The "outside experts" team recommended publication of the probability based indications themselves separate from the regular time sensitive reports (Williams, Steinberg, and Jessen 1984).

> "We support the idea of publication of the probability based indications for use by the technical community, but not necessarily at the same time as the Crop Reporting Board estimates and with reservations as indicated below. These probability based data, together with all measures of precision and accuracy should be considered as proper for public access if appropriately presented. The exact sequence of publication would need to be developed to assure continued public credibility for the CRB estimates while at the same time satisfying the technical community that the Agency is fulfilling its professional responsibility for permitting proper examination of probability based data."

The approach that the agency has taken is to provide past performance information in most major crop and livestock reports. This "track record" is included as part of a reliability writeup which includes a short description of survey procedures, sample sizes, and timing. Performance normally consists of two measures. A "root mean squared error" is calculated by expressing all deviations from the final estimates for the last 20 years as percentages of the final estimates, squaring the deviations, averaging the sum of the squares, and taking the squared roots. This root mean squared error does improve as the season progresses. For example, the corn root mean squared error for November 1 is 2.5%. The corresponding figure for August 1 is 8.5%, largely due to extreme drought in 1983 which did not occur until after August 1 and cut forecasted yield potential by 18.8%. (September root mean squared error drops to 5.1%.)

Besides listing the root mean squared error, two other depictions of reliability are included. The root mean squared error is converted to a 90% confidence interval, both as a percentage and as millions of bushels. The other measure of performance is to list the last 10-year record of differences between the current report and the final estimate. This is expressed as the average, minimum, and maximum differences and the number of years, out of 10, below and above the final.

Table 4. Reliability of March 1 hog estimates[1]

	Root Mean Squared Error	90% Confidence Level	Difference Between First and Final Estimates[2]			Number of Quarters[2]	
			Average	Smallest	Largest	First Above Final	First Below Final
	Percent		1,000 Head			Number	
All Hogs & Pigs	0.9	1.6	897	−53	−31	−1130	7	5
Pig Crop	1.2	2.2	510	67	−43	−562	5	7
Expected Farrowings								
Next Quarter	2.7	4.9	163	−17	−16	132	8	4
Following Quarter	3.8	6.7	206	27	−2	214	6	6

[1] Based on quarterly data from March 1989 through December 1991.
[2] Records for the past 12 quarters.

This "track record" approach is very useful in answering questions and record books are kept for all components of crop forecasts: planted acres, harvested acres, yield, and production. There is currently an effort to improve and standardize the reliability statements between various crop and livestock reports. Table 4 is the reliability table from the March 1992 *Hogs and Pigs* report.

7. Publication of Indications Rather Than Estimates

As stated above, most agency reports involve the interpretation of current survey indications compared to historic relationships. However, there are some major exceptions to this approach in which it is more appropriate to publish survey indications.

The most significant example is the annual farm production expenditures series. These "estimates" come from a large scale, integrated survey which collects data on farmers expenditures for the past year and collects, for subsamples of farmers, detailed data for the cost of production of specific crop and livestock commodities which rotate each year. All operators in the survey (Farm Costs and Returns Survey) receive at least aggregate production expenditure questions and operations not selected for cost of production data in the current survey answer more detailed expenditure data. The survey program is jointly funded by NASS and the Economic Research Service (ERS), which publishes the cost of production data and develops budget generators to update estimates for years when no new data are collected. ERS is also responsible for farm income estimates and uses the expenditure data in those series.

There are no check data for most expenditure items comparable to data for crops such as soybeans. Therefore, rather than developing a set of estimates for the 2600 or so possible categories in the *Production Expenditures* release, survey expansions and sampling errors are published. This may seem inconsistent with the discussion above but production expenditures data users are a very select portion of the total agricultural audience. Most users are economists in ERS and in universities and economic research centers who are developing models to estimate costs of production, expenditures, and farm income. They are nearly all familiar with the meaning of sampling errors, although they often put too much reliability in FCRS expansions with sampling errors over 20%. Published farm income figures which are more widely used are estimates, instead of indications, and no measures of sampling reliability are included.

NASS has also adopted a somewhat different procedure for the publication of survey indications for a new series on agricultural chemical usage. For the most part, these estimates will come from the agency's objective yield surveys in which every acre of a crop in a state has the same chance of selection and each interview thus has equal weight in the state estimates. Sample sizes for many states and most crops will be quite small. Sampling errors for most items are directly related to sample sizes since application of a particular chemical (fertilizer, insecticide, herbicide, etc.) becomes essentially a "yes, no" situation. (There is more variation in number of applications and rates applied.)

The users of this new data series will be extremely varied. There is also a great amount of data to present. The decision was made to not include sampling variation estimates in every table. Instead, a reliability section was included which presented percentage ranges of sampling errors for typical tables, based on sample sizes. This is expected

to be a good compromise for the desires of all users of the data.

8. Statistical Defensibility Which Might be Unpopular

As indicated, the agency has tried to be responsive to audience desires and uses of data. However, in at least one instance, NASS made a major procedural change which increased statistical defensibility but was unpopular with data users because they saw no difference in estimates but received data later.

The agency has had quarterly estimates of hogs and pigs on hand and pigs born the previous three months since 1964. Until the mid-1980s, surveys were begun the last week or 10 days of each quarter with few data collected after the first of the month. Estimates could then be published about the 24th of the month such as March 24 relating to December–February births and March 1 inventories. However, the questions asked were essentially how many hogs and pigs do you have *today* (the date of the interview, not March 1) and how many pigs were born or expected to be born in December, January, and February (although nearly all data were usually collected before February was over). A similar approach was used to collect grain stocks data before the first of the month.

Once again, the agency was using historic information on hog slaughter, exports, etc., and the approach of repeating the same survey year after year to make consistent current estimates. However, the hog industry is not stagnant; it is constantly changing towards large operations producing a greater share of total production. The collection of most data before the nominal reference data was not a statistically defensible procedure.

As part of a new integrated survey program which combined the sampling for on-farm grain stocks, hogs and pigs, and crop acreage and production, a reference date approach was developed, tested, and implemented. Data collection starts as of the first of the month and all grain stock and hog inventories are asked as "your inventory as of March 1" regardless of when the respondent is contacted. The pig crop questions relate to the previous three full months (such as December, January, and February) so there is no extrapolation required on the part of the producer.

This new reference date approach is clearly more statistically defensible and should result in more stable estimates as the industry continues to change. However, in order to allow at least 12 days for data collection and to complete all necessary editing, analysis, summarization, and estimation steps, estimates are now usually published on the last working day of the month of data collection (such as March 31). Most data users did not appreciate that the new data series might be improved; they just wanted their numbers so the change seemed like a degradation of quality to them. However, these comments and complaints soon faded and, after three years of the new timing, most data users no longer focused on the fact that numbers were once available earlier.

9. Applications to Data Confidentiality

As pointed out earlier, most data collected by NASS are on a voluntary basis. They are collected with a strict pledge of confidentiality; individual reports will not be released and will only be used for statistical purposes. The agency has long used some simple guidelines that any published total must be based on three or more reports and no one unit should make up more than 60% of a total. Exceptions to these rules require signed waivers by the producers affected.

However, there are many other defensibility factors that the agency has considered concerning unpublished data. For example, many of the potential state or regional data which could be published from the Farm Costs and Returns Survey are not published because the expansions cannot be defended. The issue is not confidentiality per se (no one could identify who the 20 operations were in the sample for the whole country who had a certain category of expense) but it is a data reasonableness issue.

In examining requests for publishing more detailed breakouts of information, NASS has tried to adopt a statistically defensible approach. How many observations were in that cell? (We might be reluctant to publish if there were 30 or fewer responses.) What was the coefficient of variation for the item requested? (If it exceeds 20%, are the estimates really helpful to the user?)

NASS has made other specific decisions concerning release of data because of what it considered to be lack of statistical defensibility. For example, brands and model numbers of machinery purchased are collected on the Farm Costs and Returns Survey in order to properly edit size and price ranges. One data user wanted a tabulation of brands purchased. This was turned down as not being reasonable information to extract from the survey.

In a second example, a decision was made, for agricultural chemical usage data, not to show zero indications for chemicals applied to crops in specific states. Since sample sizes are small (less than 100 per state in many cases) the fact that the current survey did not pick up a herbicide in one state which is used in adjoining states is more likely a function of sampling variability rather than an indication that none was used. With the wide variety of data users expected to access those tables, it was decided not to use a "not reported" or zero answer. Instead, any states without positive reports are included with states with only a few reports into an "other states" total so the summary level totals are complete.

10. Statistical Defensibility and Total Quality Management

NASS, like many other statistical organizations, has considered Total Quality Management (TQM) principles in an effort to improve its performance and products. A number of people in the agency have received TQM training and a major effort was made to apply TQM principles to a survey process control evaluation of the Agricultural Survey Program.

The agency has not adopted major changes in its operations due to TQM. Nearly everyone who has received specific training or experience in TQM has concluded that the NASS statistical defensibility and other improvement efforts had already resulted in NASS being a "TQM agency". Some of the reasons for those conclusions include:

1. The NASS functional form of organization means that multiple units (questionnaire design, survey training, sampling frame section, commodity estimation units, State Statistical Offices) are involved in the planning and execution of all survey activities.
2. The NASS approach of progressing in a person's career by taking on more responsible positions in a variety of State Statistical Office and headquarters assignments results in professional staff members who do understand the workings and issues in other offices of the agency.
3. The agency has been successful in having working groups, task forces, and ad hoc committees working across

organizational lines to solve problems.

4. Specifications for all major surveys are set by collecting input from all organizational units involved in the survey processes.
5. Suggestions for improvements are solicited and received for all survey operations.
6. Customer inputs are regularly received through scheduled data users' meetings and all-day briefing presentations in conjunction with major reports.

In order to more fully examine the role that TQM might play in improving agency performance and products, representatives from each of the units involved in the quarterly Agricultural Survey Program were asked to serve on a Survey Quality Team. Team members were given TQM training and then formed separate groups to look at various aspects of survey process control.

Survey process control, as envisioned by NASS, involves four aspects: relevance, accuracy, timeliness, and resources. The groups looked at each aspect and summarized their findings in a Baseline Quality Report (NASS 1990). The study identified additional improvements that could be made on all aspects. A number of followup activities are ongoing. The amount of post-survey analyses of reporting methods, effects of nonoverlap operations, response rates, percentage of coverage of population, variations from survey to survey for the identical strata, etc., has been expanded and is completed in time for use by the Agricultural Statistics Board. Informal teams have examined all aspects of the report preparation process in headquarters. That review and the added communications was invaluable as processing has changed from hard copy to diskettes to the present Local Area Network used for transfer of all tabular and narrative files.

11. Continuing and Future Activities

Some 13 agency activities were listed above which contribute to statistical defensibility. Work is proceeding in each of those directions and will continue. Particularly important are efforts to standardize questionnaires and instructions for data series such as prices received by farmers. Those data series are constantly being audited and interrogated since they can have a great effect on the United States Treasury. (A one cent difference in the agency's estimate of average marketing year prices farmers received for corn would normally result in about $60 million of additional or lower payments to farmers through price deficiency programs.)

The agency hopes to use even more of a data base processing approach for all steps from sample frame maintenance through sampling, editing, and estimation. This brings in new concerns about data sensitivity and utilizing defensible procedures if additional historic information might be used for editing and estimation.

The other major effort of the agency is to tie statistical defensibility concepts into efforts to utilize total quality management principles for ongoing improvements in procedures. For all these reasons, NASS does not visualize that statistical defensibility will ever be reduced to one manual or one set of procedures.

12. References

Chern, W., Hushak, L., Jordan, J., and Bullock, J.B. (1984). Report of SRS Review, July 30–August 1, 1984 by AAEA Representatives, unpublished report.

Gilmore, H.W. (1980). *Young and Detroit v. Klutznick and Barabba*, Civil Action No. 80–71330, Opinion by Judge Horace W. Gilmore, U.S. District Court, Eastern District of Michigan, Southern District.

Hildreth, C., Finkner, A., and Goldberg, S.

(1984). Statistical Reporting Service Agency Review by ASA representatives. July 30–August 1, 1984, unpublished report.

National Agricultural Statistics Service (NASS), USDA, (1990). Agricultural Survey Program: Baseline Quality Report by the Survey Quality Team for the Agricultural Survey Program. USDA, Draft report dated March 12, 1990, Administratively Confidential.

Statistical Reporting Service, USDA (1985). Crop Reporting Board Standards. Washington, DC.

Statistical Reporting Service, USDA (1983a). Framework for the Future. Washington, DC.

Statistical Reporting Service, USDA (1983b). Scope and Methods of the Statistical Reporting Service. Misc. Pub. No. 1308.

Statistical Reporting Service, USDA (1969). The Story of U.S. Agricultural Estimates. Misc. Pub. No. 1088.

Williams, S., Steinberg, J., and Jessen, R. (1984). Agency Review by Outside Experts. Meeting, July 30–August 1, 1984.

Wolter, K.M. (1982). A Note on Statistical Defensibility: Comment. The American Statistician, 36, 209–211.

Received April 1991
Revised August 1992