Gauge R&R studies that incorporate baseline information

Nathaniel T. Stevens\textsuperscript{a}, Stefan H. Steiner\textsuperscript{a}, Ryan P. Browne\textsuperscript{b} & R. Jock MacKay\textsuperscript{a}

\textsuperscript{a} Business and Industrial Statistics Research Group, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

\textsuperscript{b} Department of Mathematics and Statistics, University of Guelph, Guelph, ON, N1G 2W1, Canada

Accepted author version posted online: 10 Sep 2012. Published online: 11 Jul 2013.

To cite this article: Nathaniel T. Stevens , Stefan H. Steiner , Ryan P. Browne & R. Jock MacKay (2013) Gauge R&R studies that incorporate baseline information, IIE Transactions, 45:11, 1166-1175, DOI: 10.1080/0740817X.2012.723842

To link to this article: http://dx.doi.org/10.1080/0740817X.2012.723842

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Gauge R&R studies that incorporate baseline information

NATHANIEL T. STEVENS, STEFAN H. STEINER, RYAN P. BROWNE and R. JOCK MACKAY

1 Business and Industrial Statistics Research Group, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada
E-mail: shsteine@uwaterloo.ca
2 Department of Mathematics and Statistics, University of Guelph, Guelph, ON N1G 2W1, Canada

Received October 2011 and accepted July 2012

The standard plan for gauge reproducibility and repeatability studies is for each of \( r \) operators to measure \( k \) parts \( n \) times for a total of \( N = krn \) measurements. These studies are usually planned and conducted in isolation ignoring available baseline data generated by the measurement system used for inspection or process control. This article has two goals. First, it quantifies the substantial benefits of incorporating baseline data into the analysis of measurement study data. Second, it searches for good standard plans with a fixed total number of measurements that take into account available baseline data. With operator effects being considered to be fixed, situations where the part by operator interaction is excluded from or included in the model are investigated. The analysis of the combined data is based on maximum likelihood estimation and the ranking of plans on the approximate standard errors of the estimates using the Fisher information matrix. The benefit of incorporating baseline data into the analysis is significant and most of the gains in precision can be obtained with small baseline sample sizes. In general, depending on the context and number of baseline measurements, the standard plan with either the minimum or maximum number of parts is recommended.

Keywords: Available data, gauge reproducibility and repeatability, maximum likelihood estimation, measurement system assessment, study design, historical data

1. Introduction

Periodic assessment of important production measurement systems is mandated within many quality systems such as ISO 9001:2008 (Myhrberg, 2009) and TS 16949:2009 (Automotive Industry Action Group (AIAG), 2009). The standard Measurement System Assessment (MSA) approach, as proposed in AIAG (2010), is a gauge Reproducibility and Repeatability (R&R) study in which several operators repeatedly measure a small set of parts. Since we are interested in the periodic assessment of measurement systems that are routinely in use, we often have single measurements on many units from regular production. We call these measurements the baseline data. We can and should incorporate these freely available data into both the plan and analysis of an MSA study. The use of baseline information has been recommended in the literature but not well studied to our knowledge. Danila et al. (2008, 2010) quantified the substantial advantage of using available information in the assessment of a binary measurement system. For continuous measurements, Steiner and MacKay (2005) suggested incorporating available baseline information into the R&R analysis when there are no operator effects. However, in the analysis they assumed that the total variation (due to both the measurement and production processes) is known rather than estimated and did not quantify the effect of their proposal. Browne et al. (2009) quantified the gains in power for testing a hypothesis about the intra-class correlation in the special case when the process mean and standard deviation are known. Minitab (2008) allows users to specify a “historical standard deviation” in an R&R analysis. AIAG (2010, p. 121) also has a short section on “Using Historical Variation Information” in measurement system assessment. However, both the Minitab and AIAG suggestions ignore operator information available in the baseline data and, as in Steiner and MacKay (2005), assume the total variation is known rather than estimated. Finally, all of these previous references other than Steiner and MacKay (2005) did not suggest altering the design of a gauge R&R study when baseline data are available. The two primary goals of this article are to quantify the effect of (properly) including baseline data into the MSA study analysis for a continuous measurement characteristic and to consider the best MSA study plans when baseline data are available.

In assessing a measurement system with \( r \) operators, it is common practice to randomly sample \( k \) parts from the current process and have each operator measure each part \( n \) times for a total of \( N = krn \) measurements. For a fixed...
number of operators \( r \), we denote this plan as MSA\((k, n)\). Common choices suggested by the default forms in the AIAG manual (2010, pp. 224–225) are \( k = 10; r = 2, 3; n = 2, 3 \) so \( 40 \leq N \leq 90 \). In our experience, most MSA studies in industry follow these guidelines. If the measurement system is automated or has a single operator \( (r = 1) \), a common choice is \( k = 10, n = 6 \). In this instance, an alternative plan was proposed by Shainin (1992); see also Traver (1995), who recommended MSA\((30,2)\) where \( k = 30 \) parts are selected and each is measured twice; i.e., \( n = 2 \). Steiner and MacKay (2005) suggested MSA\((3,20)\) (with non-randomly selected parts) when baseline data are available.

Stevens et al. (2010) considered a new set of MSA plans in which a standard plan is augmented with additional measurements to increase the number of parts in the study. With the total number of measurements (including both the standard and augmented plans) fixed, they showed that the augmentation is beneficial and recommended an allocation of effort to the two components of the design. We rely heavily on the derivations in Stevens et al. (2010) since including baseline data corresponds to what they called type A augmentation. The main difference between this work and that of Stevens et al. (2010) is that here we assume that the baseline data are readily available from previous use of the measurement system. In Stevens et al. (2010), the extra data were collected as part of the MSA study. Baseline data are commonly available when assessing a measurement system in current use.

We have organized the article as follows. In the next section, we propose a model and derive the likelihood function for the MSA\((k, n)\) plan when it is augmented by baseline data with identifiable operators and apply the likelihood analysis to an example. Then in Section 3 we report the results of a simulation study to determine when the likelihood-based asymptotic results can be safely used to rank plans. More important, we use the asymptotic standard deviation of the estimator for the gauge R&R to compare various MSA\((k, n)\) plans for different baseline sample sizes and to suggest the optimal plans. We consider the no (one) operator case as well as situations involving multiple operators with and without part by operator interaction. We see a large improvement in precision for estimating the R&R even when the baseline sample size is small. Also, in most situations the recommended plans are markedly superior to those used in practice. In Section 4, we briefly look at the case when the operator making the measurement is not identifiable in the baseline data. We end with a discussion and a summary of the results.

2. Modeling and likelihood analysis of MSA\((k,n)\) with baseline data

We denote the data from a standard R&R study by \( \{y_{ijl}, i = 1, \ldots, k; j = 1, \ldots, r; l = 1, \ldots, n\} \) so that \( i \) indexes the parts, \( j \) the operators, and \( l \) the repeated measurements. We use the following model to specify the attributes of interest in the measurement system and to analyze the data collected for the assessment:

\[
Y_{ijl} = P_i + \mu_j + PO_{ij} + M_{ijl},
\]

where \( Y_{ijl} \) is a random variable representing the observed response for repeated measurement \( l \) by operator \( j \) on part \( i \). \( P_i \) is a random variable representing the variation of the true dimension of part \( i \). \( \mu_j \) represents the effect of operator \( j \). and \( M_{ijl} \) is a random variable that represents the measurement error when the same operator repeatedly measures the same part (i.e., measurement repeatability). With this notation, the relative bias between two operators \( p \) and \( q \) is given by \( \mu_p - \mu_q \). We also include the random variable \( PO_{ij} \) to allow the operator effect to change from part to part; i.e., to allow for interaction between the operators and parts.

In this article, we consider the typical manufacturing setting where it is sensible to model the operator effects \( \mu_j, j = 1, \ldots, r \) as fixed since the number of operators is small and all are included in the assessment study. Typically we have \( 1 \leq r \leq 4 \). In other contexts, a large number of operators may use a measurement system and, in that case, it makes sense to consider random operator effects since only a sample of the possible operators can be used in the study. Burdick et al. (2005) discussed the issue of fixed versus random operator effects and provided an Analysis of Variance (ANOVA)-based analysis for standard plans in both cases. They also provided a large number of references.

To specify model (1) completely, we make the additional distributional assumptions \( P_i \sim N(0, \sigma_{P}^2) \), \( PO_{ij} \sim N(0, \sigma_{PO}^2) \), \( M_{ijl} \sim N(0, \sigma_{M}^2) \) and that all of these random variables are independent. Note that the random effects are each quantified parsimoniously by a single standard deviation. We also define \( \sigma_e^2 = \sum_{j=1}^{r} (\mu_j - \bar{\mu})^2 / r \), where \( \bar{\mu} = \sum_{j=1}^{r} \mu_j / r \) as in Burdick et al. (2005, p. 83) so that \( \sigma_e^2 \) quantifies the variation due to the relative biases among the operators (i.e., reproducibility). However, note that the parameter \( \sigma^2 \) is not a variance in the usual sense. We then define the total variation as

\[
\sigma^2 = \sigma_p^2 + \sigma_o^2 + \sigma_{po}^2 + \sigma_{m}^2.
\]
gauge R&R (AIAG, 2010) is defined as
\[
\gamma = \sqrt{\frac{\sigma_0^2 + \sigma_{po}^2 + \sigma_m^2}{\sigma_i^2}}. \tag{3}
\]
According to AIAG (2010, p. 78), a measurement system is deemed to be acceptable if \(\gamma\) is less than 0.1, unacceptable if \(\gamma\) is greater than 0.3, and is in need of improvement if 0.1 < \(\gamma\) < 0.3. If the estimate of \(\gamma\) is large, we can examine the estimates of \(\sigma_0\), \(\sigma_{po}\), and \(\sigma_m\) separately to identify the sources of the large measurement system variation.

We assume that estimation of \(\gamma\) is of primary interest. The recommended plans reflect this assumption. We also define two parameters of secondary interest, \(\delta\) and \(\beta\), to partition the variation due to the measurement system, as
\[
\delta = \frac{\sigma_m^2}{\sigma_0^2 + \sigma_{po}^2 + \sigma_m^2} \quad \text{and} \quad \beta = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{po}^2}. \tag{4}
\]
Since \(\gamma\), \(\delta\), and \(\beta\) are defined as ratios, we assume, without loss of generality, that the total variation as given by Equation (2) equals one, so \(\gamma^2 = \sigma_0^2 + \sigma_{po}^2 + \sigma_m^2\). To assess plans, we look at three cases where \(\sigma_m^2 = \delta \gamma^2\) for \(\delta = 0.1, 0.5, 0.9\) so the repeatability contribution to \(\gamma\) is relatively small to large, respectively. Then, for given values of \(\gamma\) and \(\delta\), we consider further the three cases \(\beta = 0.1, 0.5, 0.9\) where the contribution of \(\sigma_0^2\) is a relatively small to large proportion of \(\sigma_0^2 + \sigma_{po}^2\), the operator contribution to the measurement system variation (repeatability). Algebraically, we have \(\sigma_0^2 = \beta (1-\delta) \gamma^2\) and \(\sigma_{po}^2 = (1-\beta)(1-\delta) \gamma^2\). Note that setting \(\beta = 1\) leads to \(\sigma_{po} = 0\); i.e., the case where there is no part by operator interaction. If we set \(\delta = 1\), there are no operator effects in the model.

The baseline data can take two forms. Usually, we assume that for operator \(j, j = 1, 2, \ldots, r\), there are single measurements on \(b_j\) different parts and that we associate the baseline measurements with a particular operator. Then the baseline data can be denoted simply as \(z_{ijl}, j = 1, \ldots, r, l = 1, \ldots, b_j\) and according to the model (1), we have the corresponding independent random variables \(Z_{ijl} \sim N(\mu_j, \sigma_p^2 + \sigma_{po}^2 + \sigma_m^2)\). Note that the baseline data gives information about the operator means \(\mu_j\) and the within-operator process variation. Alternately, in some situations we may not be able to associate the baseline measurements with a particular operator. Then the baseline data can be denoted simply as \(z_i, i = 1, \ldots, b\). Assuming that each operator measures parts with a known relative frequency the corresponding independent random variables \(Z_i\) are distributed as a (known) finite mixture of \(r\) normal distributions, each with the same standard deviation but possibly different means.

When there are data from an MSA\((k, n)\) plan with no baseline data, we can estimate \(\gamma\) using ANOVA (Burdick et al., 2005). However, when we add baseline data, we resort to likelihood methods based on the model (1), as outlined below, to produce the estimates, as it is not clear how to adapt the ANOVA analysis. Following Stevens et al. (2010), the log-likelihood contribution from data in the MSA\((k, n)\) study is
\[
l_1(\mu, \sigma_p^2, \sigma_{po}^2, \sigma_m^2) = -\frac{1}{2} \left[ a_1 \sum_{i=1}^{k} \sum_{j=1}^{r} \sum_{l=1}^{n} (y_{ijl} - \mu_j)^2 + a_2 \sum_{i=1}^{k} \sum_{j=1}^{r} \left[ \sum_{l=1}^{n} (y_{ijl} - \mu_j) \right]^2 \right.
\]
\[+ a_3 \sum_{i=1}^{k} \left[ \sum_{j=1}^{r} \sum_{l=1}^{n} (y_{ijl} - \mu_j) \right]^2 \right] - \frac{k}{2} \ln \left[ \left( \sigma_m^2 + n \sigma_{po}^2 + r \sigma_p^2 \right) \left( \sigma_m^2 + n \sigma_{po}^2 + r \sigma_p^2 \right)^{r-1} \right]. \tag{5}
\]
where \(\mu = (\mu_1, \ldots, \mu_r), a_1 = \sigma_m^{-2}, a_2 = -\sigma_{po}^2/(\sigma_m^2 + n \sigma_{po}^2), \) and \(a_3 = -\sigma_{po}^2/(\sigma_m^2 + n \sigma_{po}^2)(\sigma_m^2 + n \sigma_{po}^2 + r \sigma_p^2)\). For the baseline data, each operator measures different parts once and measurements on all parts are independent. Thus, the log-likelihood contribution from the baseline data is
\[
l_b(\mu, \sigma_p^2, \sigma_{po}^2, \sigma_m^2) = -\frac{1}{2} \sum_{j=1}^{r} \sum_{i=1}^{b_j} (z_{ijl} - \mu_j)^2 \sigma_p^2 + \sigma_{po}^2 + \sigma_m^2
\]
\[= \frac{r b}{2} \ln \left( \sigma_p^2 + \sigma_{po}^2 + \sigma_m^2 \right), \tag{6}\]
where \(b = \sum_j b_j\) the total baseline sample size. Assuming that the parts in the baseline and the MSA\((k, n)\) study are different and hence independent, the overall log-likelihood is the sum of the two log-likelihood components given by Equations (5) and (6), namely,
\[
l_1(\mu, \sigma_p^2, \sigma_{po}^2, \sigma_m^2) + l_b(\mu, \sigma_p^2, \sigma_{po}^2, \sigma_m^2). \tag{7}\]
To estimate \(\mu_j, j = 1, \ldots, r, \sigma_p^2, \sigma_{po}^2, \) and \(\sigma_m^2\) and hence \(\gamma, \delta, \) and \(\beta,\) given by Equations (3) and (4), we maximize the log-likelihood function given by Equation (7). We also use the inverse of the Fisher (expected) information matrix and the delta method to derive standard errors for the parameters of interest. See the Appendix in Stevens et al. (2010) for further details.

### 2.1. Example
In the production of aluminum pistons, an automated gauge is used for 100% inspection of the skirt diameter at a particular height as well as many other key characteristics of the piston. There is a routine assessment of the system every 6 months using an MSA\((10, 6)\) plan. There are no operator effects. The data are shown in Table 1. The gauge reports the deviation from nominal in micrometers. From process monitoring, there were 96 measurements available from the previous 24 hours of production. The mean and standard deviation for these baseline measurements are 0.56 and 2.88, respectively. With no operator effects, these are sufficient statistics for the baseline data.
In what follows, we compare plans using the asymptotic standard error for \( \gamma \), denoted by \( SE(\hat{\gamma}) \), calculated from the Fisher information matrix. To check that the asymptotic results will allow us to appropriately rank the possible MSA(\(k,n\)) plans for different baseline sizes we first conducted a simulation study. In the simulation we compared the simulated and asymptotic standard errors for a variety of plans and parameter values. We considered:

- \( N = 60, 90 \) and \( 120; \)
- number of operators \( r = 1, 2, 3, \) and \( 4; \)
- number of parts \( k = 3 \) to a maximum depending on \( N \) and \( r; \)
- per operator baseline sample sizes of \( b_j = 0, 10, 30, \) and \( 100 \) (recall that \( b = rb_j; \))
- parameter values \( \gamma = 0.1, 0.3, \delta = 0.1, 0.5, 0.9, \) and \( \beta = 0.1, 0.5, 0.9. \)

For each plan and set of parameter values, we generated 10 000 samples from model (1) and for each sample determined the maximum likelihood estimate of \( \gamma \). The results show that the asymptotic standard error for \( \gamma \) closely matches the simulated results for all plans when the baseline sample size is large. For simulations based on small baseline sample sizes, the asymptotic results underestimate the simulated results with increasing large differences for plans with fewer parts \( (k) \). Where there was a large difference between the asymptotic and simulated results, the estimate also has substantial bias.

These differences are important if we wish to select the overall number of measurements \( N \) to meet a goal in terms of the standard error of the estimate for \( \gamma \) when the baseline sample size is small. However, for fixed values of \( N, r, b, \) and the parameter values, the asymptotic and simulated standard errors provided the same ranking of plans as \( k \) and \( n \) varied. As a result, we proceed to rank plans based on the asymptotic results. Note that we never recommend a plan whose asymptotic properties do not closely match simulation results. In addition, based on simulation results there is an even larger benefit from using available baseline data than suggested in Figs. 1 to 3.

To address the issue of bias when there was no or little baseline data we tried using REstricted Maximum Likelihood (REML) estimation. Using the transformation suggested by Corbeil and Searle (1976), we can split the log-likelihood (for both the standard plan and the baseline data) into two pieces, one that depends only on the variance components \( (\sigma^2_p, \sigma^2_o, \) and \( \sigma^2_a) \) and the other that depends on all of the variance components and the operator means \( (\mu_j, j = 1, 2, \ldots, r) \). We found the REML estimates for the variance components by maximizing the first piece of the log-likelihood. For balanced plans without baseline data the REML estimates match those obtained by the usual ANOVA estimation. We subsequently obtained estimates for the operator means by maximizing the overall log-likelihood with the variance components fixed at the REML estimates. To explore the usefulness of the REML

### Table 1. Piston diameter data from MSA(10, 6)

<table>
<thead>
<tr>
<th>Part</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.8</td>
<td>-3.6</td>
<td>-1.8</td>
<td>-2.5</td>
<td>-5.5</td>
<td>-3.3</td>
</tr>
<tr>
<td>2</td>
<td>1.8</td>
<td>2.7</td>
<td>2.0</td>
<td>2.9</td>
<td>0.8</td>
<td>-0.2</td>
</tr>
<tr>
<td>3</td>
<td>-3.7</td>
<td>-3.6</td>
<td>-3.9</td>
<td>-3.2</td>
<td>-3.4</td>
<td>-3.9</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>-1.6</td>
<td>1.1</td>
<td>0.2</td>
<td>0.9</td>
<td>-0.2</td>
</tr>
<tr>
<td>5</td>
<td>2.3</td>
<td>3.6</td>
<td>1.0</td>
<td>2.0</td>
<td>5.2</td>
<td>3.4</td>
</tr>
<tr>
<td>6</td>
<td>-2.7</td>
<td>-3.0</td>
<td>-3.2</td>
<td>-1.5</td>
<td>-1.9</td>
<td>-4.7</td>
</tr>
<tr>
<td>7</td>
<td>0.9</td>
<td>2.8</td>
<td>1.2</td>
<td>0.2</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>0.9</td>
<td>0.8</td>
<td>2.1</td>
<td>-0.2</td>
<td>0.8</td>
<td>2.1</td>
</tr>
<tr>
<td>9</td>
<td>1.3</td>
<td>1.6</td>
<td>1.5</td>
<td>0.3</td>
<td>-0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>-0.3</td>
<td>-0.6</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Using maximum likelihood estimation of Equation (7) and the observed information, we have \( \hat{\gamma} = 0.354 \) with standard error 0.0424 and an approximate 95% confidence interval for \( \gamma \) of (0.271, 0.437) when we include the baseline data. If we ignore the baseline data, the ANOVA estimate is \( \hat{\gamma} = 0.408 \) with standard error 0.086, which gives the 95% confidence interval of (0.230, 0.576). We get a large improvement in the precision of the estimate of \( \gamma \) by incorporating the available baseline data into the analysis. We provide MATLAB code and instructions for its use to calculate the maximum likelihood estimate for \( \hat{\gamma} \) and the associated standard error at the website www.bisrg.uwaterloo.ca/.
approach, we conducted a factorial simulation study similar to that described earlier in this section that compared the REML and usual maximum likelihood estimates for a number of different plans, baseline sizes, and values for $\gamma$, $\delta$, and $\beta$. The results suggest that the REML estimator of $\gamma$ is indeed substantially less biased than the usual maximum likelihood estimator (though still not unbiased) when there are no baseline data, especially when the number of parts in the MSA is small. However, when we add even a small amount of baseline data, say 30 observations, the difference in bias between the two estimation approaches disappears and in some combinations of the parameter values the usual maximum likelihood estimators are less variable than the REML estimators. For this reason, and the additional complexity of the REML approach, in the rest of the article we use standard maximum likelihood estimation.

3.1. No (or single) operator case

When there are no operator effects or only one operator (i.e., $r = 1$) we have $\sigma_o = 0$, $\delta = 1$, and $\sigma_{po} = 0$. As a result, there are only three unknown parameters $\mu_1$, $\sigma_p$, and $\sigma_m$, in model (1). The parameter of interest simplifies to $\gamma = \sqrt{\sigma_m^2/(\sigma_m^2 + \sigma_p^2)}$. To illustrate the contribution of the baseline data, we use a total of $N = 60$ measurements in the MSA($k,n$) study and consider three different plans. The plan MSA(10,6) matches the AIAG (2010) recommendation, MSA(30,2) is the plan with the maximum number of parts as suggested by Shainin (1992), and MSA(3,20), the plan with the minimum number of parts, as used by Steiner and MacKay (2005) when baseline data are also available.

Figure 1 shows the asymptotic standard error of the estimate by baseline size for the three plans when $\gamma = 0.2$. The pattern of the results is similar for other values of $\gamma$ and $N$ through $\text{SE}(\hat{\gamma})$ changes.

We draw two conclusions. First, the value of including the baseline data is substantial especially for those MSA plans with few parts. This is not surprising because without the baseline data the MSA plans with few parts provide little information about $\sigma_p$. Second, the best plan depends on whether or not baseline data are available. If there are no baseline data, MSA(30,2) is the best for estimating $\gamma$ and it results in a substantial reduction in the standard error for $\hat{\gamma}$ compared with MSA(10,6), the default AIAG plan.

With the MSA(30,2) plan and $b = 0$ (i.e. no baseline data), we match the degrees of freedom available to estimate the two unknown variance components $\sigma_p^2$ and $\sigma_m^2$. However, even with a small number of baseline observations, say $b \geq 30$, the other two plans are better than the plan with the maximum number of parts and they are much better with larger baseline sample sizes.

Note that when $b$ is large and there are only two variance components, the baseline sample gives a precise estimate of the total variation $\sigma_t$, as given by Equation (2), and hence the best plan for estimating $\gamma$ is to repeatedly measure a single part. We do not recommend this plan since with it we cannot check the linearity assumptions as discussed earlier.

As a general guideline, with a single operator, if $b > N/2$, we recommend the three-part plan. Otherwise, we suggest the plan with the maximum number of parts ($n = 2$).

For a more detailed analysis, we provide software at the website www.bisrg.uwaterloo.ca/ to find optimal plans with the following inputs:

- the number of operators $r$ (here $r = 1$);
- the baseline size $b$;
- a range of values for $\gamma$;
- the maximum number of measurements $N$ used in the MSA plan so that $kn \leq N$.

For each value of $\gamma$, the output includes the number of parts, $k$, and the number of repeated measurements per part, $n$, that minimize the asymptotic standard error of the maximum likelihood estimate $\hat{\gamma}$. As well, we show the ratio of the standard errors for the optimal plan compared with the three-part plan and the plan with the maximum number of parts. For example, if we specify $r = 1$, $b = 60$, and $N = 60$ with $0.05 \leq \gamma \leq 0.40$ in increments of 0.05, we get the output as shown in Table 2.

In this case, the best plan is somewhat sensitive to the unknown value of $\gamma$ but the standard error of the recommended three-part plan is virtually identical to that of the optimal plan over the entire range of values for $\gamma$.

3.2. Two or more operators—no part by operator interaction

Now we consider the case of two or more operators with no part by operator interaction; i.e., we set $\sigma_{po} = 0$ in model

![Figure 1](image-url)
Table 2. Optimal choice of $k$ and $n$ with $r = 1$, $b = 60$, $N = 60$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$n$</th>
<th>$\gamma$</th>
<th>$SE(\hat{\gamma})$</th>
<th>Relative efficiency of:</th>
<th>Three-part plan</th>
<th>Max parts plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>20</td>
<td>0.05</td>
<td>0.0065</td>
<td></td>
<td>1.0000</td>
<td>0.8674</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.10</td>
<td>0.0129</td>
<td></td>
<td>1.0000</td>
<td>0.8682</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.15</td>
<td>0.0194</td>
<td></td>
<td>1.0000</td>
<td>0.8695</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.20</td>
<td>0.0258</td>
<td></td>
<td>1.0000</td>
<td>0.8715</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.25</td>
<td>0.0322</td>
<td></td>
<td>1.0000</td>
<td>0.8740</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>0.30</td>
<td>0.0386</td>
<td></td>
<td>0.9996</td>
<td>0.8769</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.35</td>
<td>0.0448</td>
<td></td>
<td>0.9983</td>
<td>0.8798</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.40</td>
<td>0.0510</td>
<td></td>
<td>0.9961</td>
<td>0.8827</td>
</tr>
</tbody>
</table>

Note that as $\delta$ gets closer to one, the fraction of measurement system variation attributed to the repeatability component increases. We assume that $\gamma$ is the primary parameter of interest and base comparisons of plans on the asymptotic standard error of $\hat{\gamma}$. We assume that each operator measures $b/r$ parts in the baseline so the total baseline sample size is $b$. In Fig. 2, we compare the performance of plans with $k = 3$, 10, or 30 parts for varying $b$ when $\gamma = 0.2$, $r = 2$, $N = 120$, with $\delta = 0.1$, 0.5, and 0.9. Two of the selected plans have the minimum ($k = 3$) and maximum number of parts ($n = 1$) for the given values of $r$ and $N$. The third plan is close to that recommended by AIAG (2010).

We see again the large benefit of the baseline information for the precision of the estimator of $\gamma$ in all of the plans and that much of the benefit is obtained with $b < 80$. As in the single operator case, plans with few parts benefit the most. For small values of $b$, the best plan uses the maximum number of parts and as $b$ becomes large, the best plan uses three parts. The switching point depends on $\delta$, the relative contribution of the repeatability to the overall measurement system variation. If $\delta$ is close to one, the three-part plan becomes optimal for smaller values of $b$. We see the same general pattern when we look at similar plots for varying values of $N$, $r = 2$, 3, 4, and $\gamma$. For plans with $r$ operators we suggest the three-part plan for $b > 40 + 20r$ and the plan with a maximum number of parts otherwise.

We can use the software described earlier for a more detailed analysis. Suppose we have $N = 60$, $r = 3$, and $b = 60$. According to the proposed planning guidelines, we should use the maximum parts plan MSA(20,1). From Table 3, we see that the optimal design depends on $\gamma$ and $\delta$. However, the recommended 20-part plan has very high efficiency and we would consider using another plan only if we were confident that $\delta$ was close to one. If we set $\delta = 1$,

![Fig. 2. SE(\hat{\gamma}) as a function of baseline size with N = 120, \gamma = 0.2, \ r = 2 (solid line, MSA(10,6); dashed line, MSA(30,2); dotted line, MSA(3,20)).](image-url)
1172

Stevens et al.

Table 3. Optimal choice of $k$ and $n$ with $h = 60$, $N = 60$, $r = 3$

<table>
<thead>
<tr>
<th>Optimal plan</th>
<th>Relative efficiency of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>$n$</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

there are no differences among the operators and we return to the previous case with no operator effects.

3.3. Two or more operators—possible part by operator interaction

Next we consider the case where we include the interaction term given in model (1). There are now three parameters $\gamma$, $\delta$, and $\beta$ that describe the performance of the measurement system. As $\beta$ gets closer to one, the relative contribution of the interaction to the reproducibility decreases. It would be surprising to have small values of $\beta$ where the reproducibility is dominated by the interaction. Again we base our comparisons on the estimation of the primary parameter $\gamma$ and include only plans that provide estimates of all the parameters. That is, we now require $n \geq 2$. In Fig. 3, we look at the case of two operators with $\gamma = 0.2$ and compare the three plans MSA(10,3), MSA(15,2), and MSA(3,10), each with a total of 60 measurements over a range of values for $\delta$ and $\beta$. We see substantial improvement in the precision of $\hat{\gamma}$ by including the baseline data regardless of the plan. However, in comparing the three plans, over almost the whole range of values for $\delta$, $\beta$, and all values of $b$, the plan with the maximum number of parts, MSA(15,2), is optimal and substantially better than the other plans if $\delta$ and $\beta$ are small. We see the same behavior for other values of $N$ and $2 \leq r \leq 4$. When $\delta$, $\beta$, and $b$ are large, the three-part plan becomes optimal but there is little loss in efficiency in using the plan with a maximum number of parts. For example, when $N = 120$, $r = 4$, $\delta = 0.99$, $\beta = 0.9$, and $b = 400$, the relative efficiency of MSA(15,2) compared to MSA(3,10) is at least 0.95 as $\gamma$ varies between 0.1 and 0.3. When there are multiple operators and we wish to estimate a possible part by operator interaction effect we recommend the plan with the maximum number of parts in all situations.

Fig. 3. $SE(\hat{\gamma})$ versus baseline size with $N = 60$, $r = 2$, $\gamma = 0.2$ (solid line: MSA(10,3), dashed line: MSA(15,2), dotted line: MSA(3,10)).
4. Baseline data with missing operator information

For measurement systems used by two or more operators, there may be no record of the operator in the baseline data. Here, we consider the impact of this missing information. In this case, the appropriate baseline log-likelihood is no longer given by Equation (7). If we assume equal frequency of measurements by each operator but that operator information is not recorded, the log-likelihood for the baseline data $z_i$, $i = 1, 2, \ldots, b$, is

$$
\sum_{i=1}^{b} \log \left[ \sum_{j=1}^{r} \frac{1}{2\pi r \sqrt{\sigma_p^2 + \sigma_{po}^2 + \sigma_m^2}} \exp \left( \frac{\frac{[z_i - \mu_j]^2}{2(\sigma_p^2 + \sigma_{po}^2 + \sigma_m^2)}}{2(\sigma_p^2 + \sigma_{po}^2 + \sigma_m^2)} \right) \right]. 
$$

(8)

The overall log-likelihood is then given by the sum of Equations (5) and (8). Determining the maximum likelihood estimates is not difficult; however, finding the Fisher information is complicated (since the expected values are now messy). We instead use simulation to investigate the impact of the missing operator information on $SE(\hat{\gamma})$. We ran a simulation study for the MSA(10,2) plan with three operators. We considered all combinations of $\gamma = 0.1, 0.3, 0.5, \delta = 0.1, 0.5, 0.9$, and $\beta = 0.1, 0.5, 0.9$, as defined by Equations (3) and (4), and for each of the 27 combinations we used 1000 runs. In each run we simulated data for the specified plan from model (1) with parameters given by Equations (3) and (4) and estimated $\gamma$ using both the likelihood given by Equation (7) for the known operators case and the sum of Equations (5) and (8) for the unknown operator case.

Table 4 shows the loss of precision for estimating $\gamma$ for the MSA(10,2) plan with three operators and $\gamma = 0.3$. The loss of information from not knowing the baseline operators is surprisingly small. The differences in precision are larger when $\delta$ is small and $\beta$ is large; i.e., when the variation due to operators is relatively large and little of that variation comes from the part by operator interaction. We note that if $\beta$ is small so that the interaction effect is relatively large, there is little loss of precision (ratio is one). We surmise that when the interaction effect is a relatively large part of the overall operator effects, a single measurement by a known operator provides no information about the interaction. From the complete simulation results (not shown), when $\gamma$ is less than 0.3, the loss of precision is smaller than shown in Table 4. In all cases, there is little loss as the baseline size changes. We propose that, given the small differences in precision when we have baseline data without knowing the associated operators, we continue to use the planning guidance provided for the case of known operators.

The gauge R&R package in Minitab (2008) allows the user to enter a “historical standard deviation.” Similarly, the AIAG manual (2010, p. 121) discusses “Using Historical Variation Information.” The implicit assumption in these two approaches is that we know $\sigma_i$, as defined in Equation (2), as it is calculated as the sample standard deviation from a large historical data set. However, the estimate can be seriously biased if there are large operator effects and the data are not appropriately weighted with respect to the operators. Neither AIAG nor Minitab recommends changing the standard R&R plan in light of the available baseline information. Only the analysis is changed. There is also no mention of the fact that we may or may not have additional information about the operator effects $\mu_j$, $j = 1, \ldots, r$.

With two or more operators, if the historical standard deviation is calculated by operator and we also know the operator effects, then we use the recommendations corresponding to large values of $b$ to select an appropriate MSA study. For the case of no part by operator interaction, we recommend the three-part plan. We can investigate various plans over a range of values for $\gamma$, $\delta$, and $\beta$ if we suspect that there is interaction. We can estimate the remaining unknown parameters using the likelihood based on the selected MSA plan.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\delta$</th>
<th>$\beta$</th>
<th>$SE(\hat{\gamma})$ with known baseline operators</th>
<th>$SE(\hat{\gamma})$ with unknown baseline operators</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.1</td>
<td>0.1</td>
<td>0.058</td>
<td>0.058</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0.1</td>
<td>0.5</td>
<td>0.049</td>
<td>0.051</td>
<td>1.05</td>
</tr>
<tr>
<td>30</td>
<td>0.1</td>
<td>0.9</td>
<td>0.039</td>
<td>0.042</td>
<td>1.08</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.1</td>
<td>0.050</td>
<td>0.050</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.5</td>
<td>0.046</td>
<td>0.047</td>
<td>1.02</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.9</td>
<td>0.042</td>
<td>0.044</td>
<td>1.05</td>
</tr>
<tr>
<td>30</td>
<td>0.9</td>
<td>0.1</td>
<td>0.046</td>
<td>0.046</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0.9</td>
<td>0.5</td>
<td>0.045</td>
<td>0.045</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0.9</td>
<td>0.9</td>
<td>0.047</td>
<td>0.048</td>
<td>1.02</td>
</tr>
<tr>
<td>90</td>
<td>0.1</td>
<td>0.1</td>
<td>0.049</td>
<td>0.049</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>0.1</td>
<td>0.5</td>
<td>0.038</td>
<td>0.039</td>
<td>1.04</td>
</tr>
<tr>
<td>90</td>
<td>0.1</td>
<td>0.9</td>
<td>0.027</td>
<td>0.029</td>
<td>1.08</td>
</tr>
<tr>
<td>90</td>
<td>0.5</td>
<td>0.1</td>
<td>0.042</td>
<td>0.042</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>0.5</td>
<td>0.5</td>
<td>0.037</td>
<td>0.038</td>
<td>1.03</td>
</tr>
<tr>
<td>90</td>
<td>0.5</td>
<td>0.9</td>
<td>0.032</td>
<td>0.034</td>
<td>1.06</td>
</tr>
<tr>
<td>90</td>
<td>0.9</td>
<td>0.1</td>
<td>0.038</td>
<td>0.038</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>0.9</td>
<td>0.5</td>
<td>0.036</td>
<td>0.036</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>0.9</td>
<td>0.9</td>
<td>0.036</td>
<td>0.036</td>
<td>1</td>
</tr>
</tbody>
</table>
5. Discussion and conclusions

In this article, we address the planning and analysis of measurement system assessment studies when baseline data are available. This is a common situation for measurement systems used for inspection or process control, where the baseline data are available at no additional cost. We quantify the benefits of incorporating baseline data (parts measured once) into the measurement system assessment study analysis and show that substantial improvements in precision are possible and attained even with small baseline sample sizes. We also recommend changes to the usual MSA study plan in terms of the number of parts. With the total number of measurements \( N = kn \) fixed, we recommend MSA plans that use either the minimum or maximum number of parts unlike the industry standard plan that suggests \( k = 10 \) parts. To summarize, incorporating the baseline data into the analysis and selecting either the recommended maximum or minimum number of parts plan dramatically increases the precision of the estimate of \( \gamma \), the gauge R&R.

We see most of the benefit from incorporating the baseline data in the analysis with total baseline sample sizes as small as 60. We also found that knowing the operators used in the baseline study is helpful but only makes a noticeable difference if the operator-to-operator differences are large. To avoid bias we need to be careful that the baseline data reflect both the current manufacturing and measurement processes. In the analysis, we assumed that the process and measurement systems are stable for the time interval that covers both the collection of the baseline data and the MSA study. This suggests a question of “how much baseline data we should use?” In particular, how far back in time should we go? If the measurement system changes at some point during the collection of the baseline data, then the estimates of \( \mu_1, \ldots, \mu_r \) and \( \sigma^2 \) from the baseline will be biased. To address this potential problem, we suggest checking for stability in the baseline data as recommended in AIAG (2010).

With a single operator and a large baseline sample (say, larger than 50 observations), we saw in Fig. 1 that the three-part plan is the best. With multiple operators, if \( \delta \) (the proportion of measurement variability due to repeatability) approaches one and \( b \) (the baseline sample size) is large, then the three-part plan is optimal but the gain in precision for estimating \( \gamma \) is smaller (Fig. 3). This occurs because we are still estimating the other parameters using the full model (1). If we first collapse the model by setting \( \delta = 1 \), then we get the full benefit of the three-part plan. Of course, there is no need to use multiple operators in the MSA in this case.

If the measurement system has a single operator or no operator effects, the likelihood analysis is unnecessary because the three parts provide relatively little extra information about \( \sigma_p \) and we can estimate \( \sigma_m \) using ANOVA with little loss of precision. This idea was proposed by Steiner and MacKay (2005), who, in the situation where there is a single (or no) operator, recommend an MSA study with a large, medium, and small part (as defined by the baseline data) rather than parts selected at random from the process. In this case, the estimation approach suggested in this article is not applicable. Instead, we use the MSA data only to estimate \( \sigma_m \). A more efficient (but complicated) maximum likelihood estimation procedure for this case that takes into account the selection process is given by Browne et al. (2009) and extended in Browne et al. (2010) to situations involving multiple operators. This series of papers shows the major benefit of selecting extreme parts for the MSA study in terms of the precision of the estimate of \( \gamma \).

We have concentrated on the estimation of the primary parameter \( \gamma \). If interest centers on measurement system characteristics such as the precision to tolerance ratio (AIAG 2010), here defined as \( 6\sqrt{\sigma_p^2 + \sigma_m^2}/(USL - LSL) \) where \( USL \) and \( LSL \) are the upper and lower specification limits, respectively, that do not involve the process variation \( \sigma_p \), there is little value in the baseline data.

A referee raised the issue of whether or not the part by operator interaction should be included in the model. In the context where the measurement system is assessed periodically, we recommend including the interaction term in the initial assessments. If the estimate of \( \beta \) is close to one, then we can choose a plan and model in subsequent assessments that does not include the interaction term. In the subsequent assessments, using the parts with repeated measurements, we can informally assess a possible interaction by plotting the repeated measurements against part number with different symbols for each operator.

We have modeled operators as fixed effects, as is likely reasonable in most industrial contexts. In other situations, it may be more reasonable to assume that operators are random effects. MSA studies with random operators but without baseline data were considered by Steiner et al. (2011). To model the operators as random effects and incorporate baseline data we could consider a similar analysis to that proposed in this article. If we further assume that the operators used in the baseline and MSA study are different (and randomly selected) we can write down the likelihood and determine the Fisher information matrix. With the same operators used in the MSA and the baseline studies, the likelihood expression is complicated and finding the Fisher information matrix is difficult. Another possible extension that we do not explore further here is the possibility of replacing the standard plans with plans that are not balanced in terms of number of measurements made by each operator. An unbalanced plan may be better, for instance, when operators are not balanced in the baseline data.

References


**Biographies**

Nathaniel T. Stevens is a Ph.D. candidate in the Statistics and Actuarial Science Department and member of the Business and Industrial Statistics Research Group at the University of Waterloo.

Stefan H. Steiner is a Professor in the Statistics and Actuarial Science Department and Director of the Business and Industrial Statistics Research Group at the University of Waterloo.

Ryan P. Browne is an Assistant Professor in the Mathematics and Statistics Department at the University of Guelph.

R. Jock MacKay (retired) is an Adjunct Professor in the Statistics and Actuarial Science Department at the University of Waterloo.