Free Ridership: Arbitrary Algorithms vs. Consistent Calculations

Jeff Erickson and Mary Klos, Summit Blue Consulting, Verona, WI,
Valy Goepfrich, WPPI Energy, Sun Prairie, WI

ABSTRACT

This paper compares two different approaches to calculating free ridership from a common population of participants to examine whether changes in the questions and algorithm make a substantial difference in the final result. The study was based on a telephone survey for a custom commercial and industrial energy efficiency program. The survey included questions that supported two well-established algorithms: Summit Blue and Focus on Energy (Focus). These two approaches produced similar net-to-gross ratios, 65 percent and 60 percent respectively. The Summit Blue algorithm uses weights at two stages and the results were similar under a wide range of weight scenarios. When the Summit Blue algorithm was restricted to one of its three perspectives that was similar to the Focus algorithm in terms of the issues addressed, it produced a net-to-gross ratio of 74 percent compared with Focus' 60 percent. Taken all together the results of this study suggest that well thought-through variations in free rider algorithms can provide consistent results, but there is a need for continued research on the design of algorithms and source questions to improve their robustness.

Introduction

WPPI Energy is a regional power company serving 51 customer-owned electric utilities in Wisconsin, Upper Michigan, and Iowa. As they ramp up their energy efficiency programs, they face the challenge of ensuring that programs are achieving electricity savings that would not otherwise have occurred. In addition, from a power supply planning perspective, they face the challenge of adjusting the energy and demand forecasts for electricity savings above historical levels.

An impact evaluation of WPPI Energy's energy efficiency programs produced verified gross and net savings for several programs, with special emphasis on estimating net savings for the custom C&I programs. Free ridership was based on participant surveys – referred to as the self-report approach. The self-report approach is common for many types of programs, including custom C&I. While there is significant agreement in a broad sense that the self-report approach should address the quantity, efficiency, and timing of measures that would have been implemented in the absence of a program, there is less agreement on other areas, including:

- How much does the algorithm that is used to convert survey answers into free ridership scores matter?
- How much does the specific phrasing of the questions matter?
- If there are weights, how much do they matter?

These questions were explored in the process of estimating net savings for WPPI Energy's custom C&I programs. This paper discusses the approach taken, the results, and the implications for free ridership research.

Methodology

To explore the choice of free ridership algorithm, the phrasing of questions that provide the data for the algorithm, and weighting in the algorithm, free ridership for WPPI Energy's custom C&I programs was estimated using two well-established algorithms: Summit Blue and Focus on Energy (Focus). The Summit Blue algorithm has been used in the evaluation of programs in several other jurisdictions and it is particularly appropriate for WPPI Energy's custom C&I program. The Focus algorithm has been used for several years in the evaluation of Wisconsin's Focus on Energy business programs.[1] The Summit Blue and Focus algorithms provide an opportunity to examine the choice of algorithm, questions, and weights.

A telephone questionnaire was developed to collect data for both the Summit Blue and Focus free ridership algorithms. The questionnaire was administered to a random sample of 100 participants that made an energy efficiency improvement in 2007 and received financial or technical assistance from WPPI Energy's custom C&I program. Data necessary for both the Summit Blue and Focus algorithms were collected from all respondents. In many cases, the two algorithms used data from the same question. Although, as we discuss later, in a few instances it was necessary to collect data using two different questions, one suited for the Summit Blue algorithm and another suited for the Focus algorithm.

The Summit Blue algorithm produces a free ridership score for each sampled project. The Focus algorithm produces a free ridership score for each sampled measure. In each case, a ratio estimator is used to combine these individual free ridership scores to produce program-level attribution (i.e., a net-to-gross ratio). For example, in the case of the Summit Blue algorithm, the ratio estimator uses project-level program attribution based on the project-level free ridership scores, project-level evaluated gross electricity savings, and the sample weights.

The Summit Blue algorithm addresses free ridership from three different perspectives: program influence, measure-level, and best estimate. The Focus algorithm considers free ridership from only the measure-level perspective, and in a manner similar but not identical to the Summit Blue algorithm. In addition, the timing and efficiency questions are phrased differently. Next, we discuss each of the algorithms in more detail.

Summit Blue Algorithm

The Summit Blue algorithm addresses free ridership from the three perspectives summarized earlier: program influence, measure level, and best estimate. It calculates a free ridership score from each of the perspectives and then combines the three free ridership scores to obtain the final free ridership score. Each of these perspectives is discussed separately below (see Figures 1, 2, and 3).

Program influence. The program influence free ridership score (see Figure 1 and Table 1) is based on several program influence questions and a question regarding plans prior to the customer's initial contact with the program. These questions are designed to clarify the role that the program played in the decision to make the improvement and to provide supporting information on free ridership:

1 WPPI's programs in Wisconsin operate outside of the statewide Focus on Energy portfolio of programs.

- Point BB in Figure 1 represents the maximum of the importance scores (1 to 5) assigned by the respondent to the financial assistance, the technical assistance, and their relationship with the utility.

- Point DD is a score (1 to 5) assigned by Summit Blue staff based on the response to a closed (yes/no) and an open-ended question (if "yes") regarding the influence of the program on the type, efficiency level, or quantity of the improvement.

- Point FF is a score assigned by Summit Blue staff based on the response to a closed (yes/no) and an open-ended question (if "yes") regarding plans in place prior to the customer's initial contact with the program. This score ranges from 1 for documented plans and budget to 5 for no plans.

The three values (BB, DD, FF) are averaged (GG) and then converted to the program influence free ridership score (HH) assuming a linear relationship between the score (1 to 5) and free ridership (0 to 100 percent). In the example shown in Table 1, the answers produce a program influence free ridership score of 17 percent.

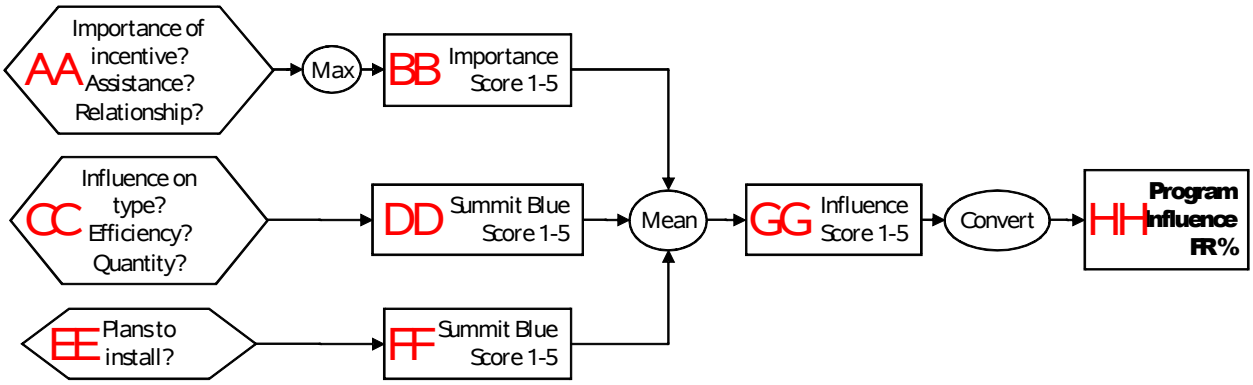

Figure 1. Summit Blue Free Ridership Algorithm – Program Influence

Table 1. Summit Blue Free Ridership Algorithm – Program Influence Example Calculation

Figure 1	Question	Possible Values	Example	Calculation
AA	How important was the financial assistance in your decision to install ___?	1 = not at all important; 5 = very important	4	
AA	How important was the technical assistance in your decision to install ___?	1 = not at all important; 5 = very important	2	
AA	How important was your ongoing relationship in your decision to install ___?	1 = not at all important; 5 = very important	1	
BB	Maximum importance	1 = not at all important; 5 = very important	4	Max(AA)
CC-DD	How did the assistance you received influence the type, efficiency, or amount you installed?	1 = program had no influence; 5 = program was the primary reason that energy efficient equipment was installed	4	
EE-FF	Describe any plans you had to install the equipment prior to receiving assistance	1 Documented plans/budget; 5 no plans	5	
GG	Average three scores	1-5	4.3	MEAN(BB,DD,FF)
HH	Convert program influence score to free rider percentage assuming a linear relationship	0-100%	17%	

Shaded cells are calculated values.

Measure-level. The measure-level free ridership score is based on questions asked separately for each measure (see Figure 2 and Table 2). If a respondent would not have made any changes to the equipment addressed by the improvement without the program in the foreseeable future (point A in

Figure 2), their free ridership score for that measure is zero (point B). On the other hand, if the respondent would have done something, they are asked about the timing, efficiency, and quantity of the improvement without the program. These three data elements are combined to produce the measure-level free ridership score (point I) as described next.

The answer (in months) to the timing of the improvement without the program determines the early replacement multiplier shown at point C in Figure 2:

- If the improvement would have been made within a year, the multiplier is 1 (that is no adjustment).
- If the improvement would have been made between one and two years the multiplier is 70 percent.
- If the improvement would have been made between two and four years the multiplier is 35 percent.
- If the improvement would have been made more than four years later the multiplier is zero.

The early replacement multiplier recognizes that for some part of the improvement's life, the free ridership score may be zero. Rather than maintain different free ridership scores for different years of an improvement's life, we reduce the measure-level free ridership score. For measures that would have been installed in the distant future (4 years or more in this case), the measure-level free ridership score is multiplied by an adjustment factor of zero because it is assumed that there is too much uncertainty regarding what a customer thinks they may do more than four years in the future. In jurisdictions where the free rider value should be calculated solely based on first year savings, this adjustment multiplier would not be applicable.

Points D and E in Figure 2 represent the answer to the following question "Without the assistance provided by your utility and WPPI Energy, in terms of percentage how likely is it that this improvement would have been at the same high level of efficiency? Please provide a percent between 0 and 100 representing the likelihood." If the improvement involved a single unit, the free ridership percentage at point H equals this likelihood at point E.

If the improvement involved more than one unit, the respondent was also asked "Without the assistance provided by your utility and WPPI Energy, what share of the equipment for this improvement would you have installed anyway at the same high level of energy efficiency?" (points F and G in Figure 2). In the case of multiple units, the free ridership percentage at point H is the product of the likelihood (E) and share (G) percentages. By allowing respondents to think about both the efficiency and quantity that would have been installed in the absence of the program, this approach allows the respondents to give an answer that most applies to their situation.

It has been argued that multiplying factors in a free ridership algorithm incorrectly lowers the free ridership result because of compounding effects (Keating). The Summit Blue algorithm includes multiplication at two points. If there are two factors and either is 100 percent, the free ridership result is unaffected by this type of compounding error. In the Summit Blue algorithm, this is the case if without the program either the improvement would have been at the same high level of efficiency with 100 percent likelihood or the same quantity (100 percent) would have been installed at the same high level of efficiency. On the other hand, if both these factors are less than 100 percent, then multiplying them produces a measure-level free ridership score that is smaller than it would be based on either of the two factors alone. We believe the compounding effect is appropriate in the case of applying these two factors to the free-ridership estimate.

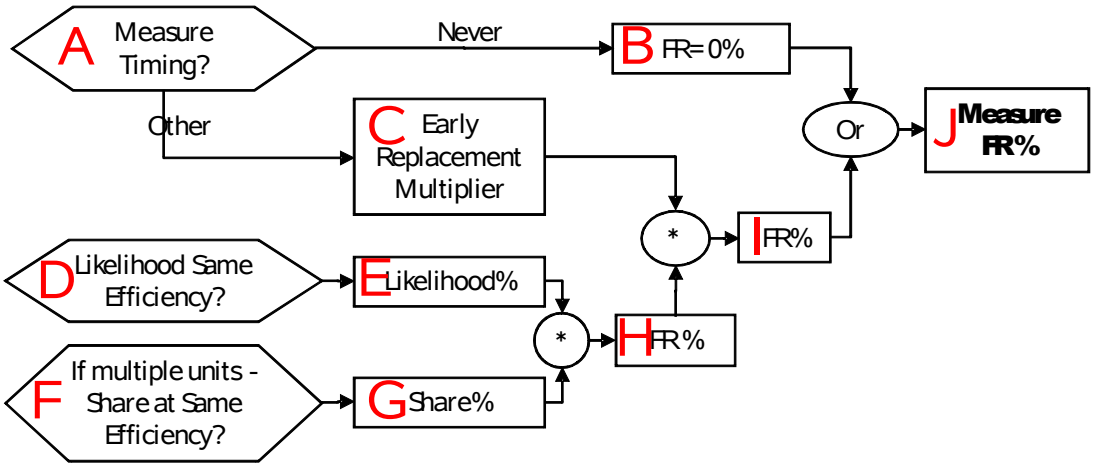

Figure 2. Summit Blue Free Ridership Algorithm – Measure Level

Table 2. Summit Blue Free Ridership Algorithm – Measure Level Example Calculation

Figure 2	Question	Possible Values	Example	Calculation
A	Without assistance from your utility and WPPI, would your organization have [installed measure] last year or in the foreseeable future? (Yes/No)	Yes, No	Yes	
A	If yes, When?	Number of months	5	
C	Early replacement multiplier	0-100%	100%	<=12 months=100% else <=24 months=70% else <=48 months=35% else 0%
D-E	Without the assistance provided by your utility and WPPI, in terms of percentage how likely is it that this improvement would have been at the same high level of efficiency?	0-100%	75%	
F-G	Without the assistance provided by your utility and WPPI, what share of the equipment for this improvement would you have installed anyway at the same high level of energy efficiency?	0-100%	50%	
H	Combine Likelihood and Share		38%	E*G
I	Adjust by early replacement multiplier (C)	0-100%	38%	C*H
J	Measure level free ridership score	0-100%	38%	If (A=No) then Zero else use value at I

Shaded cells are calculated values.

If you hold the efficiency constant but lower the quantity that would have been installed without the program, then it is logical that the free rider rate should decrease. Similarly if you hold the quantity constant but lower the efficiency that would have been installed without the program, then it is also logical that the free rider rate should decrease. Since these two decisions are made independently of each other, it seems appropriate that if both the quantity and efficiency would have been lower without the program then the free rider rate should be lower than either of the other two values alone. This effect is illustrated in Figure 3 where we see that if the program has had relatively little effect on either quantity or efficiency (both near 100%) then the product of the two is still relatively high. Similarly, if the program had a large effect on both (near 0%) then the product of the two is still relatively low. When the program has had some influence on both quantity and efficiency (say 55%) then the product of the two is substantially less than either value (30% in this example), reflecting the fact that both are at play.

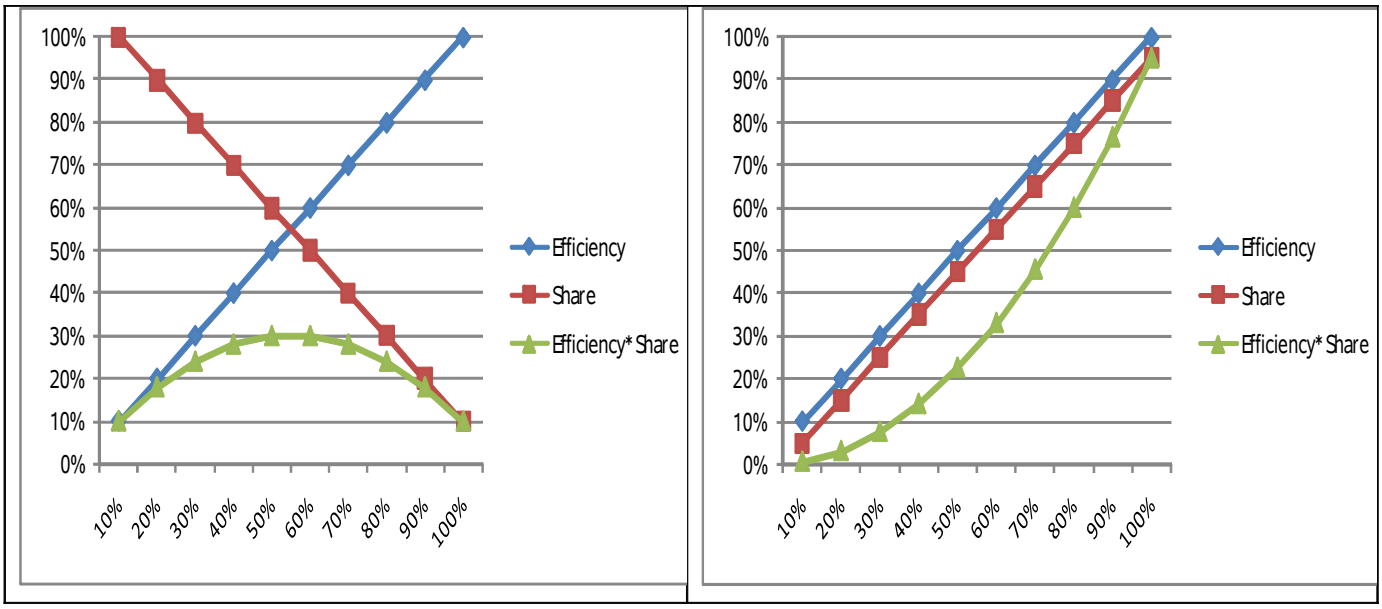

Figure 3. Effect of multiplying factors

The free ridership score should reflect both the level of efficiency and quantity that would have been installed without the program. Consider an improvement that produces 500 kWh of evaluated gross savings annually. Suppose from an efficiency perspective 20 percent of the savings (500 kWh × 20 percent = 100 kWh) would have been achieved without the program. Suppose from a quantity perspective 40 percent of the 100 kWh savings would have been achieved without the program or 40 kWh (100 kWh × 40 percent = 40 kWh). This leaves 460 kWh of savings (500 kWh – 40 kWh) attributable to the program. The result is the same if we multiply the efficiency and quantity factors to obtain the free ridership score (20 percent × 40 percent = 8 percent), subtract the free ridership score from 100 percent to obtain program attribution (100 percent – 8 percent = 92 percent), and then multiply program attribution and evaluated gross savings (500 kWh × 92 percent = 460 kWh).

Best estimate. The "best estimate" free ridership score is based on a single question addressing electricity savings as a whole: "Overall, across all the energy efficiency improvements we've been discussing, what share of the electricity savings would your organization have achieved anyway, without the assistance provided by your utility and WPPI Energy?"

Combining the 3 free ridership scores. The best estimate free ridership score at point L in Figure 4 is averaged with the measure-level free ridership score at point J to create the direct free ridership score at point M. The WPPI Energy base results weight the best estimate free ridership score and measure-level free ridership score equally in this calculation. The direct free ridership score at point M is averaged with the program influence free ridership score at point HH to create the final free ridership score at point O. The WPPI Energy base results weight the direct free ridership score and program influence free ridership score equally in this calculation. As a result of the weights, the best estimate free ridership score and measure-level free ridership score each have a weight of 25 percent in the calculation of the final free ridership score and the program influence free ridership score makes up the remaining 50 percent.

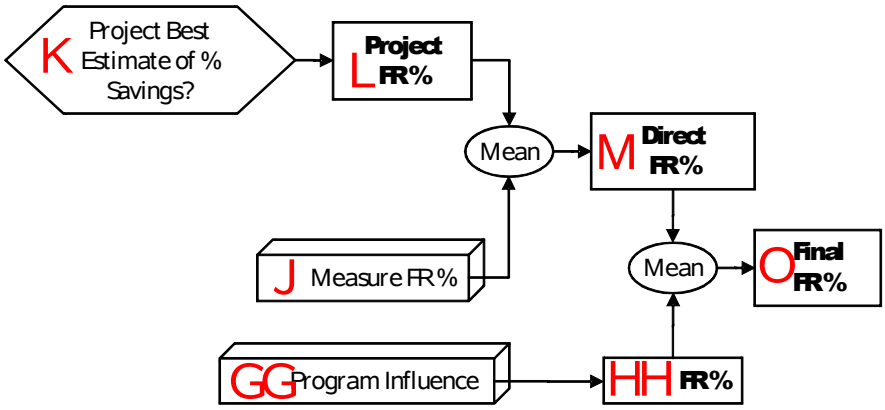

Figure 4. Summit Blue Free Ridership Algorithm – Best Estimate and Final

Table 3. Summit Blue Free Ridership Algorithm – Best Estimate and Final Example Calculation

Figure 4	Question	Possible Values	Example	Calculation
HH	Convert program influence score to free rider percentage assuming a linear relationship	0-100%	17%	
J	Measure level free ridership score	0-100%	38%	If (A=No) then Zero else use value at I
K-L	Overall, across all the energy efficiency improvements we've been discussing, what share of the electricity savings would your organization have achieved anyway, without the assistance provided?	0-100%	25%	
M	Direct free ridership score	0-100%	31%	Mean(J,L)
O	**Final free ridership score**	0-100%	**24%**	Mean(HH,M)

Shaded cells are calculated values.

Focus Algorithm

The Focus algorithm (see Figure 5) considers free ridership from only one perspective, which is similar to the Summit Blue measure-level free ridership score. Focus calculates this free ridership score in a manner similar but not identical to Summit Blue, and a couple of the questions are phrased differently as shown in Table 5. In addition to the differences between the initial and efficiency questions, while both the Focus and Summit Blue algorithms include an early replacement adjustment, the adjustments are different.

The respondent is initially asked the likelihood that they would have made the improvement without the program (fA in Figure 5). This is the first point where the Focus method uses a different question than the Summit Blue method (see Table 5 for the difference in question wording). Where Summit Blue initially asks a yes/no question on whether or not the respondent would have installed the same measure in the same time frame without the program, Focus allows for a probabilistic response to the same basic question. Those that respond 'Very Unlikely,' 'Don't Know,' 'Refused,' or 'No Answer' to the initial Focus question are asked additional probing questions to aid in the expert assessment of their free ridership score (fB and fC in Figure 5). In comparison, respondents who answer 'no' to the initial Summit Blue question are automatically assigned a free ridership score of zero.

If the respondent replies to the initial Focus question that there was some likelihood that they would have made similar improvements on their own, they are asked about the efficiency, quantity, and timing of the improvement they would have made without the program. These three data elements are

combined to produce the Focus free ridership score (fK in Figure 5) as described next.

Assessing planned efficiency improvements in the absence of the program is the second area where the Summit Blue and Focus methods use different survey questions (see Table 5). While the Summit Blue method asks for the respondent's assessment of the probability that they would have installed the same efficiency level on their own, the Focus method probes for the efficiency level that would have been installed using customer-friendly language. These questions represent two different ways to get at a piece of information that is difficult for respondents to quantify.

The "Efficiency %" (fE in Figure 5) for the Focus method reflects the percentage shown in Table 4 that is associated with the response to the efficiency question.

If the improvement involved a single unit, the initial free ridership score (fJ in Figure 5) equals this efficiency percentage. If the improvement involved more than one unit, the initial free ridership score is the product of the "Efficiency %" and "Share %" (fJ in Figure 5). The same quantity question provides data for both the Focus and Summit Blue algorithms.

Both the Focus and Summit Blue free ridership algorithms include early replacement adjustments, which are based on data from the same timing question, but the adjustments are slightly different based on the way the question responses are translated into free ridership adjustments. Focus adjusts a customer's initial free ridership score for early replacement by subtracting some portion of that score based on the timing question. The Focus early replacement adjustment is as follows:

- If the improvement would have been made at the same time, there is no adjustment and a customer's initial free ridership score is their final free ridership score.
- If the improvement would have been made more than four years later (or at exactly four years), a customer's initial free ridership score is reduced by 100 percent and their final free ridership score is zero.
- Otherwise, a customer's initial free ridership score is reduced by the following:
  the customer's initial free ridership score
  × (when (in months) the improvement would have been made without the program)/48.

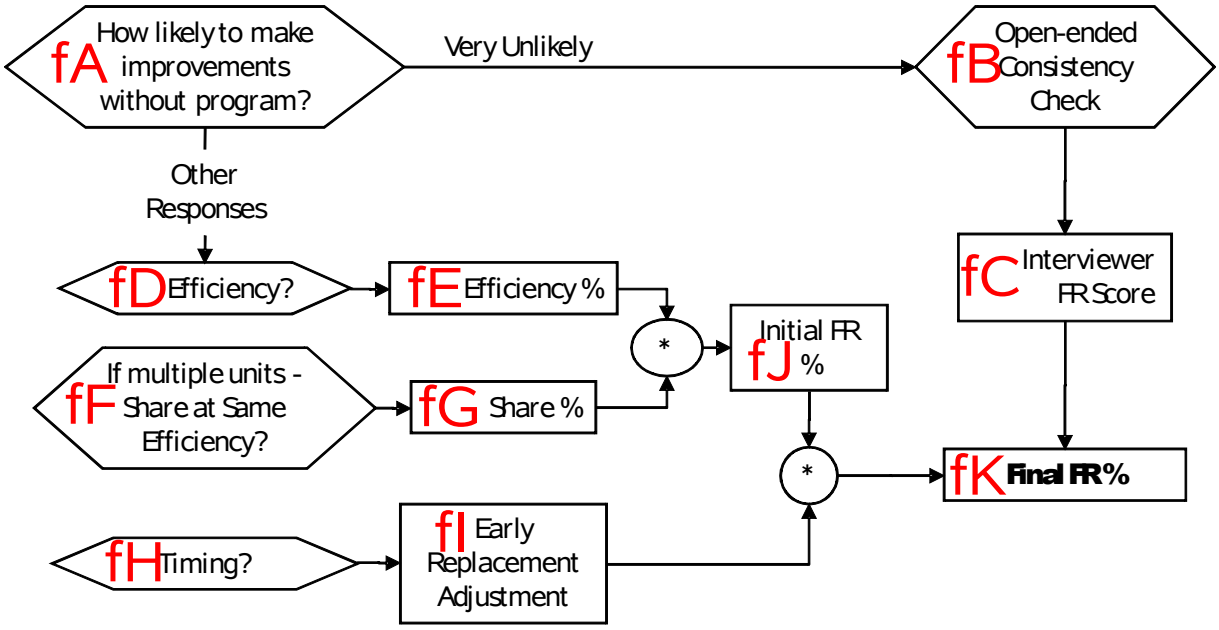

Figure 5. Focus on Energy Free Ridership Algorithm

Table 4. Focus on Energy Free Ridership Algorithm

Figure 4	Question	Possible Values	Example	Calculation
fA	Without the assistance provided by your utility and WPPI, how likely would your organization have been to [install measure]?	1 = Very unlikely; 2 = Not very likely; 3 = Somewhat likely; 4 = Very likely	2	
fB	Did the assistance you receive from your utility and WPPI in any way influence the type, efficiency level or amount of high efficiency equipment you installed? In what ways . . . (record verbatim)	Verbatim comments	NA	Only used if fA = 1 (Very Unlikely)
fC	Expert coding of verbatim response	1 = No Program Influence = 100% FR; to 5 = Program was Primary Influence = 0%	NA	Only used if fA = 1 (Very Unlikely)
Figure 3 D-E	Without the assistance provided by your utility and WPPI, in terms of percentage how likely is it that this improvement would have been at the same high level of efficiency?	0-100%	25%	From Summit Blue question battery; Ignored unless it was 100%; if 100% then fE = 100%
fD	Without the assistance provided by your utility and WPPI, would you say the efficiency would have been . . .	1 = Std Effic or according to code 2 = Slightly higher than std effic 3 = Midway between std effic and the high effic that was used 4 = Slightly lower than the high effic that was used	2	
fE	Efficiency Score - FR %	0-100%	30%	1 = 10%, 2 = 30% 3 = 50%, 4 = 70%
fF-fG	Without the assistance provided by your utility and WPPI, what share of the equipment for this improvement would you have installed anyway at the same high level of energy efficiency?	0-100%	50%	
fJ	Initial FR % before adjustment for timing	0-100%	15%	fE*fG
fH	Without assistance from your utility and WPPI, would your organization have [installed measure] last year or in the foreseeable future? (Yes/No) If yes, When?	Number of months	36	
fI	Early Replacement Adjustment		25%	1 - (months/48); If months >48 then = 0%
fK	Final FR %	0-100%	4%	fJ*fI; if fA = 1 (Very Unlikely) then fK = fC

Table 5. Differences in Focus vs. Summit Blue Questions

Question	Summit Blue	Focus
Initial	Without assistance from your utility and WPPI Energy, would your organization have [e.g., increased the energy efficiency of the compressed air system] last year or in the foreseeable future? • Yes • No	Without the assistance provided by your utility and WPPI Energy, how likely would your organization have been to [e.g., increase the energy efficiency of the compressed air system]? Would you say you would have been: • Very likely • Somewhat likely • Not very likely • Very unlikely
Efficiency	E2a Without the assistance provided by your utility and WPPI energy, how likely is it that this improvement would have been at the same high level of efficiency? • 0 to 100 percent	[SKIP if E2a=100%] Without the assistance provided by your utility and WPPI Energy, would you say the efficiency would have been • Standard efficiency or according to code (10% FR) • Slightly higher than standard efficiency (30% FR) • Midway between standard efficiency and the high efficiency that was used (50% FR) • Slightly lower than the high efficiency that was used (70% FR)

Results

Summit Blue vs. Focus

The Summit Blue and Focus free ridership algorithms produced similar net-to-gross ratios, 65 percent and 60 percent respectively (see Table 6). The difference of 5 percentage points is relatively small, but it is statistically significant (10 percent level).

Table 6. Comparison of Summit Blue vs. Focus Results

Free Ridership Algorithm	Net-to-gross Ratio	90% C&I
Summit Blue	65%	61% - 69%
Focus	60%	53% - 67%
Summit Blue measure-level only	74%	68% - 80%

As discussed, the Focus algorithm considers free ridership from only one perspective, which is similar to the Summit Blue measure-level free ridership score. In the case of the Summit Blue algorithm, if instead of basing the net-to-gross ratio on the free ridership scores obtained from the complete algorithm, we base it on only the measure-level free ridership scores, the net-to-gross ratio is 74 percent compared with Focus' 60 percent. The 14 percentage point difference is statistically significant (10 percent level).

Although the net-to-gross ratio based on Summit Blue's measure-level free ridership scores only was not the ultimate net-to-gross ratio, it is instructive to explore what is causing the difference between the Summit Blue measure-level free ridership score and the Focus free ridership score.

Table 7 compares the Summit Blue and Focus results (sample weighted) for the various components of Summit Blue's measure-level free ridership score and Focus' free ridership score. This comparison shows there was very little difference in the contribution of the quantity adjustments and the timing adjustments to the overall estimates of free ridership. The main areas of difference come from the initial question and the efficiency free ridership fraction. This makes sense since these are the areas of major difference between the two algorithms where different questions were used to gather data.

Between these two, the greatest difference comes from the initial question. In the Summit Blue algorithm, 38% of respondents choose the 'no' response, saying that they would not have installed the measure without the program. This immediately gives these customers a measure-level free ridership score of zero percent. In the Focus algorithm, no one is automatically assigned a free ridership score of zero percent. Even those who say they are 'very unlikely' to have installed the same measure without the program still get asked probing questions to develop some estimate of free ridership.

Table 7. Areas of Difference for Summit Blue Measure-level Only vs. Focus

Component	Summit Blue	Focus
Initial question: % of customers assigned FR=0%	38%	None
Efficiency FR fraction	68% likelihood of same efficiency	42% of installed efficiency level

	level	
Quantity adjustment (Share %)	52%	52%
Timing adjustment	78%	79%

This difference in the two methods at the measure level is justifiable. The Summit Blue method can make an extreme assumption of zero free ridership at this point because they know there are additional areas of inquiry that will be followed and the extreme assumption will be averaged with other values that test its validity before a final free ridership estimate is made. The Focus method must use more discretion at this point because this is their only method. They must add validity to their assumptions at this point because this will be the final answer.

Given these differences, it seems that the Summit Blue final answer would be expected to be the most comparable to the Focus final answer. It is comforting to see the similarity between the two estimates of free ridership (NTG 65% to NTG 60%) from this perspective.

Weights in the Summit Blue Algorithm

The Summit Blue free ridership algorithm combines results from three perspectives: program influence, measure-level, and best estimate . The net-to-gross ratio of 65 percent reflects weights of 50 percent/50 percent at each stage: the combination of measure-level and best estimate free ridership scores, and the combination of the resulting direct free ridership score with the program influence free ridership score. The sensitivity of the results to the weights was examined at each stage using weights of 100 percent/0 percent, 75 percent/25 percent, 25 percent/75 percent, and 0 percent/100 percent. The point estimate of the net-to-gross ratio under each of these weight scenarios (see Table 8) falls within the 90 percent confidence interval of the net-to-gross ratio using weights of 50 percent/50 percent (90% CI = 61% - 69%). Given that a reasonable argument can be made for a variety of weight scenarios, the fact that the results are similar under a wide range of scenarios is a beneficial characteristic of the Summit Blue algorithm.

Table 8. Summit Blue Algorithm: Weights Scenarios

Stage 1 Direct Free Ridership Score					
Measure-level weight	100%	75%	50%	25%	0%
Best estimate weight	0%	25%	50%	75%	100%
Net-to-gross ratio	69%	67%	65%	63%	61%
Stage 2 Final Free Ridership Score					
Direct weight	100%	75%	50%	25%	0%
Program influence weight	0%	25%	50%	75%	100%
Net-to-gross ratio	66%	66%	65%	64%	64%

Conclusions

The difficulty in measuring free ridership accurately has been used by some to argue for abandoning the field altogether. However, others argue that a reasonable estimate of free ridership is better than no estimate. WPPI Energy, for example, uses free ridership results to improve their energy

efficiency programs and net savings estimates to adjust their electricity forecast for savings above historical levels. Of course, robust free ridership estimates would be ideal.

This study provides some evidence that different free ridership algorithms can produce similar net-to-gross ratios. The Summit Blue and Focus free ridership algorithms produced net-to-gross ratios of 65 percent and 60 percent respectively. In addition, the Summit Blue algorithm uses weights at two stages and the results were similar under a wide range of weight scenarios. These results argue in favor of free ridership as consistent calculations vs. arbitrary algorithms.

On the other hand, restricting the Summit Blue algorithm to one of its three perspectives, the measure level, which is similar to the Focus algorithm in terms of the issues addressed (efficiency, quantity, and timing), produced a net-to-gross ratio of 74 percent compared with Focus' 60 percent. While it was expected that the comparison of measure-level estimate to measure-level estimate would have greater similarity, upon closer examination it became apparent that this is not an apples-to-apples comparison. The source of the difference appears to be Summit Blue's willingness to make more extreme (and simple) assumptions at this stage because they know there are additional areas of inquiry that will be followed and the extreme assumption will be averaged with other values that test its validity before a final free ridership estimate is made. The Focus method must use more discretion at this point because this is their only method. In reality, the Summit Blue final estimate is the best comparison to the Focus final estimate.

These results suggest that free ridership algorithms which use multiple lines of questioning to develop careful and thoughtful balancing of responses from a variety of angles will produce similar final free ridership estimates, and arbitrary weighting schemes used within components of the overall estimate will not create large bias in the results. Since this study only compared two alternative methods, there is a need for continued research on the design of algorithms and source questions to improve their robustness. More comparative studies like this would help demonstrate the consequences of various approaches on the robustness of the free ridership and net-to-gross estimates.

References

Keating, Kenneth M. "Freeridership Borscht: Don't Salt the Soup." *Proceedings of the International Energy Program Evaluation Conference,* Portland, Oregon. August 12, 2009.

Klos, Mary, Jeff Erickson, Argene McDowell of Summit Blue Consulting, and George Penn of Global Energy Options. *Impact Evaluation Of WPPI's 2007 Energy Efficiency Programs.* Wisconsin Public Power, Inc. December 15, 2008.

PA Consulting. *Focus on Energy Statewide Evaluation, Business Programs Comprehensive Report: Volume III, Impact Evaluation*, Final. December 23, 2002.