Data Source Assessment and Recommendation Process
Prepared and Submitted by UNAHA

Initial Screening

This step is designed to prevent wasting time researching and documenting the characteristics of nominated sources that clearly will not work in the IHBG formula (e.g., the nominated data collection project is a repackaging of Census data and not an independent data source). Each identified technical support person will answer the following questions for all nominated data sources. The answers will be compiled into a format similar to the matrix below and, along with any additional narrative component of the answers, will be provided to the Study Group, who will meet to reach consensus over which nominated sources do not meet these minimal requirements and should be either eliminated or re-nominated as a model.

1. Is it an independent, verifiable data source or a repackaging/special tabulation of some other data?
   <if not independent, stop and consider the source it is based on instead>
2. Is this data collection project active?
   <if the source is no longer being collected and cannot be reliably enhanced to bring current, reject or re-nominate as a model>
3. Does this source measure some aspect of Indian Housing need? What aspect(s)?
   <if the data source does not include any data relevant to Indian housing need, reject>
4. Is the project national in scope, collecting data and estimating values for all Indian areas?
   <if not currently or potentially national, reject or re-nominate as a model>
5. Does the data collection project contact and identify the eligible population?
   <if the data source does not include estimates for eligible populations, reject or re-nominate as a model>
6. Recommendation: (accept / re-nominate as model / reject) Please explain.

Example Summary Matrix to Accompany Narrative Answers:

		Technical Support 1 Recommendation	**Technical Support 2 Recommendation**	**Technical Support 3 Recommendation**
Data Source 1	Q1	yes	yes	yes
	Q2	yes	yes	yes
	Q3	yes	yes	yes
	Q4	yes	yes	yes
	Q5	yes	yes	yes
		accept	accept	accept
Data Source 2	Q1	yes	yes	yes
	Q2	no	no	no
	Q3	yes	yes	yes
	Q4	no	no	no
	Q5	no	no	no
		re-nominate as model	reject	re-nominate as model

<u>**Characterize the Data Source**</u>

These questions are designed to gather the facts that will form the basis for the judgments made during the evaluation process. Prior to passing judgment on the quality of a data source and its suitability for use in the IHBG formula during the evaluation phase, you must first gather information about it. Unlike the evaluation questions that follow, there are no predetermined right or wrong answers for any of these questions. The data sources will be divided evenly and randomly among the technical support person(s) for an initial two -week period to answer the following characterization questions, including citations. After those two weeks (or potentially longer, if there are substantial barriers to accessing the necessary information), the data sources and completed work will be redistributed among all members of the technical support team for the following two weeks. During this time, all members of the technical support team will have the opportunity to add to the narrative answers and make their own recommendation about whether the data source should move on to the evaluation stage. The full narrative, including the recommendation from each technical support person, will be distributed to the Data Study Group.

Purpose and Methodology

1. Who collects the data and for what purpose(s)?
2. What aspects of Indian housing need, including current IHBG formula variables, can this source measure?
3. What questions are used to collect the data? Please attach a copy of questionnaires and/or forms and any associated instructions/training materials and definitions.
4. For what population(s) or sub-population(s) is the data collection program designed to estimate data values?
5. For what geographic levels(s) is the program designed to estimate data values? What, if any, strategies are used to ensure suitable coverage of small and/or rural tribal areas?
6. How are the individuals or units chosen to participate (i.e., what is the sampling strategy)? Are there any segments of the eligible population not being reached?
7. How often is data collected? Is the data collected at a single point in time sample or as a rolling sample? What time period does the data reflect?
8. What procedures (for example follow up visits, incentives, marketing, etc.) are in place to encourage participation and completeness of the dataset?

Accuracy and Precision

1. What is the confidence limit used to calculate the published margin of error? If no confidence limits or margins of error are provided, confirm there was no sampling or extrapolation involved.
2. What methods are in place to deal with total and partial nonresponse among the individuals recording this data? What are the rates of total and partial nonresponse?
3. Is the relative margin of error consistent across all tribes/tribal areas (e.g., small, large, rural, urban, off-reservation, on-reservation, etc.)? If not, describe the variation.
4. Overall, what design issues (e.g., phrasing of questions, incentives for participating, imputation methods, number of attempts to collect data for each selected participant, perceived conflicts of interest, etc.) could introduce biases for all or a certain subgroups of tribes (e.g., small, large, rural, urban, etc.) or certain types of data (e.g., financial, population, etc.)?

Implementation and Funding

1. What organization(s) (e.g., HUD, tribes, TDHE, IHBG program) are responsible for implementing and administering data collection and/or analysis (including recruiting, hiring, training, and monitoring field staff, supplying necessary equipment, and compiling the results)?
2. How much do the data collection and analysis phases cost, and how are they funded? If there is a specific cost to IHBG recipients or to the IHBG allocation, specify that cost.
3. How long after collection will it take for the data to be aggregated and available for use?

Transparency and Potential for Challenge

1. For which of the above questions was it difficult or impossible to find an answer? What prevented answering those questions?
2. What procedures would be recommended for a tribe/TDHE to challenge inaccurate data from this source?
3. Could the data collection procedures be modified to deal with future modifications of the formula and/or formula areas? How? What opportunities exist to improve the accuracy and/or precision of the data source?
4. How has the data collection methodology changed over the last few data collection cycles?
5. How stable has the data been over the last few data collection cycles?

Other Potential Concerns

1. What other factors not addressed above could impact the suitability of this data source for use the IHBG formula? In what way(s)?

Recommendation

1. Should this data source move on to the evaluation stage? Why or why not?

<u>**Evaluate the Data Source**</u>
Fundamentally, in order to be appropriate for use, the data needs to be **relevant** (i.e., measure housing need according to the requirements of the NAHASDA Statute), **current** (i.e., reflect timely conditions to form the basis of yearly allocations), **accurate and precise** (i.e., reflect conditions within a known and acceptable margin of error), **complete** (i.e., include consistent information for all tribes/tribal areas), and **available** (i.e., without insurmountable financial or other barriers to collection, access, and use). Additionally, in order to assess these qualities, the data and data collection methodology needs to be **transparent**.

For a theoretically perfect data source, answers to all of the questions below would be YES. However, a simple yes or no answer to these questions would not only often be very difficult to provide without substantial qualifiers, it would be of very limited to the study group's process of weighing all the necessary factors to select the most appropriate data source(s). Each question should therefore be answered in a narrative format that identifies the ways the data source does and does not meet the criteria, with the potential recommend methods to mitigate any identified problems. The evaluator(s) will also review the merits of and potential ways to improve the data source for each the six broad categories and provide an overall summary of the strengths and weaknesses of the data source with regard to the IHBG formula.

The data sources will be divided evenly and randomly among the technical support person(s) for an initial two week period to answer the following evaluation questions. After those two weeks, the data sources and completed work will be redistributed among all members of the technical support team for the following two weeks. During this time, all members of the technical support team will have the opportunity to add to the narrative answers and make their own recommendations and judgments. To the extent possible, technical support person(s) will harmonize and reconcile their answers to provide a consistent opinion to the Data Study Group and compile their findings into a summary matrix in a similar format to the one at the end of this section. The full narrative answers to the questions, including a discussion of topics where the technical support team was unable to reach an agreement and a summary matrix for each data source will be provided to the Data Study Group. After meeting with the group and having the opportunity to request more information as necessary, the technical support persons will work together to write a preliminary and then final evaluation report.

Relevance

1. Does the data collection program contact and identify the population eligible for NAHASDA programs?
2. Are the definitions of terms used in all questions consistent with NAHASDA statutory requirements? What terms might be problematic?
3. Does the data measure some aspect of Indian housing need? Which aspects?
   <select from checklist of current and potential variables here>

Overall, is the data source RELEVANT? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source relevancy be improved? What resources are needed to make these improvements?

Currency

1. Can the data be updated yearly without artificial aging? If not, how often is the data updated?

2. Is the aggregated data available for use within a reasonable time frame after it is collected? How long does it take for the data to be available? Explain any delays.
3. Is the data stable over time? (i.e. no sudden swings in values, caused by sampling/methodology changes/etc)

Overall, is the data source CURRENT? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source currency be improved? What resources needed to make these improvements?

Accuracy and Precision

1. Does the data collection program methodology support deriving estimates covering the desired geographies (formula areas) and target population within those areas?
2. Are the margins of error reasonable and consistent across all tribal areas? Identify any major concerns about precision.
3. If ageing of the data will be necessary to incorporate the data in the formula every year, will any error introduced from the ageing process be reasonable and consistent across all tribal areas?
4. Are the questions phrased clearly to ensure respondents do not misunderstand what is being asked?
5. Are there sufficient protocols to make sure the data is collected in a culturally sensitive manner?
6. Are there sufficient protocols in place to verify the accuracy of collected data?
7. Are imputation and/or weighting methods appropriate and unlikely to introduce biases among all or a certain subgroup of tribes (e.g., small, large, rural, urban, etc.) or certain types of data (e.g., financial, population, etc.)?
8. Is it possible to correct or compensate for any and all survey design issues (e.g., phrasing of questions, incentives for participating, imputation methods, number of attempts to collect data at sampled housing unit, etc.) likely to introduce biases for all or a certain subgroups of tribes (e.g., small, large, rural, urban, etc.) or certain types of data (e.g., financial, population, etc.)?

Overall, is the data source ACCURATE and PRECISE? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source accuracy and precision be improved? What resources are needed to make these improvements?

Completeness

1. Does the program collect data for tribal areas in a uniform manner?
2. Are outreach efforts to encourage participation in the survey appropriate and effective within tribes/tribal areas? Are those efforts equally effective and equally implemented across all tribal areas?
3. Are all populations well represented in the data source, as evidenced by high response and inclusion rates or any other criteria? Explain any identified areas, populations, and/or topics where response or inclusion rates may be a cause for concern.

Overall, is the data source COMPLETE? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source completeness be improved? What resources are needed to make these improvements?

Availability

1. Can the data be collected and analyzed with no significant additional resources?
2. Is there a source of funding available for the data collection and analysis? Explain.
3. If tribes/TDHEs will be responsible for any of the process, will there be support to prevent any undue administrative and/or financial burden on all or a certain subgroup of tribes/TDHEs (e.g., small, large, rural, urban, etc.)? Explain.
4. Is the data quantifiable and easily integrated into a funding allocation formula?

Overall, is the data source AVAILABLE? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source availability be improved? What resources are needed to make these improvements?

Transparency

1. Are previously conducted studies/evaluations of the data source available?
2. Were you able to find answers to all data screening, characterization and evaluation questions?

Overall, is the data source TRANSPARENT? Choose one of the following options and explain: Excellent, Good, Fair, Poor How can the data source transparency be improved? What resources are needed to make these improvements?

Summary and Conclusions

1. What, if any, other information is necessary to fully evaluate this data source for use in the IBHG formula?
2. Overall, is the data source appropriate for the IHBG formula? Choose one of the following options and explain: Excellent, Good, Fair, Poor What are the areas of biggest concern?

Example Summary Matrix to Accompany Narrative Answers:

		Judgment	Comments and Concerns
Relevance	Technical Support 1	Excellent	
	Technical Support 2	Good	
	Technical Support 3	Excellent	
Currency	Technical Support 1	Fair	
	Technical Support 2	Fair	
	Technical Support 3	Fair	
Accuracy/ Precision	Technical Support 1	Good	
	Technical Support 2	Good	
	Technical Support 3	Good	
Completeness	Technical Support 1	Excellent	
	Technical Support 2	Excellent	
	Technical Support 3	Excellent	
Availability	Technical Support 1	Poor	
	Technical Support 2	Poor	
	Technical Support 3	Poor	
Transparency	Technical Support 1	Good	
	Technical Support 2	Good	
	Technical Support 3	Good	
Overall	Technical Support 1	Good	
	Technical Support 2	Poor	
	Technical Support 3	Good	