

26 July 2012

To: Ofelia Eugenio and Christian Hainzl, UNDP/MYA

From: Mark Schreiner

Re: Main results from pilot of the poverty scorecard

This memo discusses the main results of the pilot of the poverty scorecard with 197 households. In both Kyaiklat (Ayeyarwady) and Pindaya (South Shan), one village was censused, and another was sampled. Two villages were sampled in Hpa-an (Kayin).

### **Interviews/enumerator/day**

Across pilot sites, interviews consistently lasted about 10 minutes. This does not count time to get to the village and back, to plan a route among households, to travel between households, and to explain the survey to the community at a group meeting (or to explain it to an individual respondent). Counting everything, the all-in completion rate was about 10 interviews/enumerator/day (30 minutes per household) in Kyaiklat and Pindaya and about 15 interviews/enumerator/day (20 minutes per household) in Hpa-an. To err on the safe side, planning should assume 10 interviews/enumerator/day.

### **Estimated poverty rates**

For the national poverty line and 150% of the national line, Figure 1 shows estimated poverty rates from scoring in the six pilot villages and for the pilot overall. As benchmarks, Figure 1 also shows the IHLCA-based poverty rates for Myanmar as a whole, rural Myanmar, and rural areas in the pilot states. Except for the censused villages, the pilot sample is not representative of anything, so these figures just serve to illustrate the kind of analysis that can be done with scoring.

Subject to this caveat about non-representativeness, these six UNDP villages have about average poverty rates for their areas. The poverty rate for the national line in rural Myanmar is 25 percent, and the overall poverty rate among the 197 pilot households is 26 percent. There is thus no evidence that these six villages were selected for UNDP programs because of they had high rates of poverty.

**Figure 1: Estimated poverty rates from scoring in pilot villages ( $n = 197$ )**

Region/Township	Census?	# HHs	Estimated poverty rate (%)	
			Natl. pov. line	150% of natl. pov. line
Ayeyarwady (Kyaiklat)	Yes	56	32	81
	No	18	20	71
	All	74	29	78
Rural Ayeyarwady			29	79
Kayin (Hpa-an)	No (village 1)	33	27	77
	No (village 2)	20	21	74
	All	53	25	76
Rural Kayin			13	64
Shan (Pindaya)	Yes	50	23	72
	No	20	27	78
	All	70	24	74
Rural Shan (South)			25	76
Overall UNDP sample		197	26	76
Union of Myanmar			21	68
Rural Myanmar			25	74

Of course, UNDP does not directly serve entire villages. Livelihood assistance, for example, is targeted to particular households who, by the UNDP’s wealth-ranking process, fall in classes D or E. Of 197 households scored, 71 were D or E.

Figure 2 shows scoring’s estimated poverty rates for these 71 households. For the national line, their poverty rate is 41 percent (versus 25 percent for rural Myanmar). That is, DE households targeted via wealth ranking have a higher poverty rate (based on scoring’s consumption-based definition of poverty) than the village as a whole.

Because wealth ranking and poverty scoring use different definitions of poverty, this analysis merely shows that wealth ranking and scoring agree more often than not. It does not, of course, indicate whether targeting via wealth ranking (or via scoring) is “good enough”, nor whether one approach is better than the other.

This example with pilot data shows some basic uses for scoring’s estimated poverty rates. In particular, UNDP can report poverty rates for villages (or higher administrative levels) and compare them with benchmarks. Village poverty rates—whether from a census or a representative sample—can be used for targeting. That is, UNDP might exit from villages with low poverty rates, and enter (or remain in) villages with high poverty rates. Or it might allocate its budget across villages (or township offices) in proportion to the number of poor people being served there.

Figure 3 shows scoring’s estimated poverty rates for the 71 households with the lowest scores. (Two-thirds of these are DE households by wealth ranking; see next section.) For the national line, their poverty rate is 49 percent (versus 25 percent for rural Myanmar). This is 8 percentage points higher than for the 71 DE households, but not too much should be read into this, since the two approaches use different definitions, and the scorecard will be more accurate for its definition, and vice versa for wealth ranking.

Figures 2 and 3 also highlight that not all targeted households are poor as defined by consumption and the national poverty line. Likewise, not all DE households are poor (except by definition in the wealth-ranking system), although the extent of inaccuracy for wealth ranking cannot be known.

**Figure 2: Estimated poverty rates from scoring in pilot villages for UNDP participating households in classes D or E according to Wealth Ranking ( $n = 71$ )**

Region/Township	Census?	# HHs	Estimated poverty rate (%)	
			Natl. pov. line	150% of natl. pov. line
Ayeyarwady (Kyaiklat)	Yes	4	44	88
	No	29	38	93
	All	33	38	92
Rural Ayeyarwady			29	79
Kayin (Hpa-an)	No (village 1)	14	53	92
	No (village 2)	4	37	91
	All	18	49	92
Rural Kayin			13	64
Shan (Pindaya)	Yes	11	46	90
	No	9	31	82
	All	20	39	86
Rural Shan (South)			25	76
Overall UNDP sample		71	41	90
Union of Myanmar			21	68
Rural Myanmar			25	74

**Figure 3: Estimated poverty rates from scoring in pilot villages for the lowest-scoring 71 pilot households**

Region/Township	Census?	# HHs	Estimated poverty rate (%)	
			Natl. pov. line	150% of natl. pov. line
Ayeyarwady (Kyaiklat)	Yes	4	49	94
	No	28	44	93
	All	32	45	93
Rural Ayeyarwady			29	79
Kayin (Hpa-an)	No (village 1)	14	59	96
	No (village 2)	4	48	94
	All	18	56	95
Rural Kayin			13	64
Shan (Pindaya)	Yes	11	53	90
	No	9	43	82
	All	20	48	86
Rural Shan (South)			25	76
Overall UNDP sample		70	49	92
Union of Myanmar			21	68
Rural Myanmar			25	74

## **Concordance between targeting by scoring and by wealth ranking**

Scoring and wealth ranking define *poverty* differently. A complete, precise comparison is impossible, as wealth ranking's definition is qualitative, implicit, and varies by village. Nevertheless, it is useful to check how much the two approaches overlap when targeting. If overlap is high, then households and field workers will readily accept scoring. If overlap is low, then they will resist/resent scoring.

Figure 3 shows the overlap, with a focus on the bottom-right matrix for all 197 pilot households. The targeting cut-off for scoring is set so as to target the same number of households (71) as are now targeted as DE by wealth ranking.

Scoring targets 47 of the 71 DE households, and it does not target 24. Thus, scoring and wealth ranking overlap for two out of three targeted households.

For non-targeted households (wealth classes A, B, and C), overlap is greater: 102 of 126 cases, or four out of five. Most scorecards (and wealth ranking) separate the rich from the poor better than they separate the almost-poor from the just-barely poor.

Overall, targeting by the two approaches overlaps for 149 of 197 households, or about three out of four households.

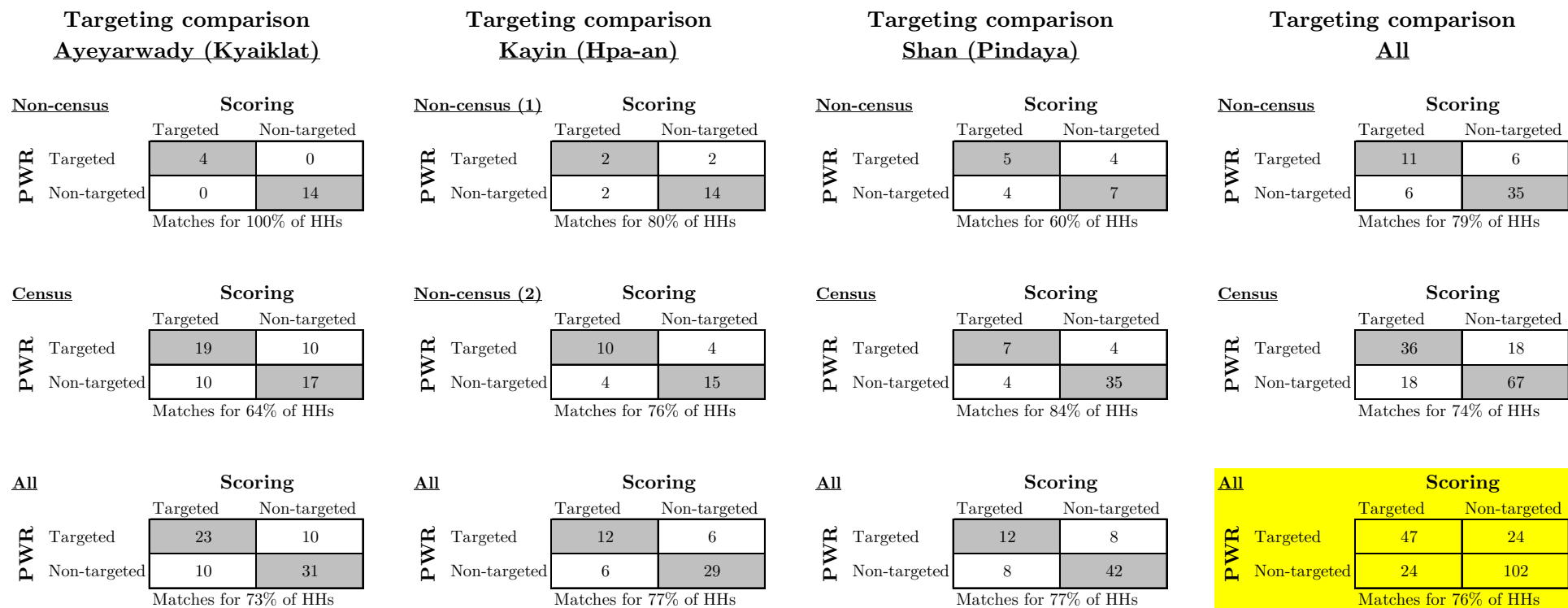
Is this overlap “enough”? My judgment is that it is a lot of overlap, but probably not enough inspire widespread buy-in by field staff. And getting buy-in is the central challenge in the process of organizational change that is adopting poverty scoring.

Of course, neither scoring nor wealth ranking are “right” nor “wrong”. They define poverty differently, and each has its own strengths and weaknesses. But wealth ranking is the incumbent, so field staff take it—by default—as the defending champion, and poverty scoring as the challenger.

UNDP/Myanmar has used wealth ranking for years, and it enjoys widespread acceptance by households and field staff, if only because it is all they know. It uses participatory processes—a central goal of UNDP—and it uses the definition of *poverty* that is used by the people affected by it. Furthermore, wealth ranking can consider more dimensions of poverty than can scoring.

Yet governments and donors use scoring's consumption-based definition. It is consistent and quantitative, allowing comparisons across villages, organizations, and time. It is the best one-dimensional indicator of poverty, highly—albeit imperfectly—correlated with nearly every other dimension of human well-being.

**Figure 4: Targeting concordance, scoring and wealth ranking**



So the one-in-four mismatch between scoring and wealth ranking is not due to “errors” by either approach. They define poverty differently, so overlap is necessarily imperfect. Nevertheless, field staff naturally consider wealth ranking to be the standard, and they believe it is more accurate than scoring. Of course, any judgment of accuracy must be relative to some benchmark, and there is no objective, universal benchmark.

All that said, it seems to me that wealth ranking is usually more accurate for targeting households (although it costs more). After all, it considers information that scoring—limited to ten indicators—ignores. For example, scoring does not distinguish between a destitute widow(er) living alone versus a high-income urban professional living alone. Scoring does not ask about health or disability, but villagers are acutely aware of health and handicaps. Scoring can ask about acres of land, but not quality of land.

In the few pilot cases that I saw with large discrepancies between scoring and wealth ranking, my subjective judgment (and that of enumerators and the other pilot observers) fit better with wealth ranking. In these cases, there were specific, identifiable reasons why the scorecard seemed off. For example, a pair of households with disabled adults were targeted by wealth ranking but not by scoring. Likewise, a one person household headed by an elderly widow (and a two person-household consisting of an elderly brother and sister) were targeted by wealth ranking but not by scoring. Wealth ranking considered—reasonably, in my view—disability and age, but scoring did not.

Scoring also targeted a couple of households that wealth ranking did not, mostly because they had many members working in agriculture. In the IHLCA data, this is associated with greater poverty risk (probably because larger, poorer households can and do send more members to work in the fields). For some large households, however, having many agricultural workers leads to greater income and higher consumption (at least according to the enumerators and the villagers). A scorecard might pick this up with an indicator for the dependency ratio, but such an indicator does not work if the scorecard already includes household size (which this one does), and, in any case, computing a dependency ratio is too complex for a simple scorecard.

Of course, most discrepancies between scoring and wealth ranking in the pilot were small and did not have clear-cut explanations. The point here is that many of the biggest mismatches had an explanation, and it almost always favored wealth ranking.

In existing UNDP villages, field agents will struggle to accept when a household that qualified for livelihood assistance under wealth ranking does not qualify under poverty scoring. (In new UNDP villages, the point is moot, and scoring can be used exclusively there without bothering staff.) When mismatches occur, field staff believe strongly that wealth ranking should trump scoring.



Thus, UNDP should not automatically expect to discard wealth ranking and replace it with scoring. Instead, it should weigh the following two options, always keeping in mind the (sometimes conflicting) goals of maintaining the good faith of field staff, keeping costs down, and targeting accurately:

- Discard wealth ranking and replace it with scoring. This is the least-costly choice. If chosen, then there must be a system in place for requesting, evaluating, and monitoring exceptions. In this way, field staff (and poor households) have recourse when scoring's targeting is off. The current system for appealing classifications by wealth ranking could be adopted to manage exceptions to scoring
- Target with wealth ranking, and estimate poverty rates with scoring. This is costs more, but it plays to each approach's strengths and avoids their weaknesses. Scoring provides objectivity and consistency across regions, summarizing poverty in a single number for management and reporting. Wealth ranking provides participation and a local, comprehensive definition of *poverty* that is already well-accepted

I, the developer of poverty scoring, hereby state directly that scoring may not be UNDP's best option for targeting households. UNDP has a useful new tool, but it may still have some uses for its old one. Wealth ranking is well-accepted, and it makes fewer big targeting mistakes. UNDP should not underestimate the capacity of field staff to subvert/undermine scoring if they resent it or do not believe in it. At the same time, wealth ranking and scoring costs more than just scoring, and UNDP will use scoring in any case to estimate poverty rates for management and reporting.

As usual, the decision boils down to weighing expected costs versus benefits. Is the benefit of avoiding unhappy field staff (and some big targeting mistakes) worth the cost of using both poverty scoring and wealth ranking? Or is the cost of mistargeting a couple of people out of ten—and of auditing field staff to discourage “cooking” data when they think scoring is off, and of managing exceptions—low enough to permit the sole use of scoring? The choice hinges on value judgments that only UNDP can make.

## Weaknesses of indicators and response options

The pilot sought to detect indicators and responses that had weaknesses such as :

- Difficult to collect
- Difficult to interpret
- Embarrassing to ask or answer
- Confusing
- Irrelevant
- Often lied about
- Time-consuming
- Has a relationship with poverty that varies by region

Related to this, the pilot sought to discover issues of interpretation so that they could be addressed in training or in the “Guidelines to the Interpretation of Indicators”. Issues were resolved by logically applying known criteria from the “Guidelines” or by asking the IHLCA Technical Committee whether enumerators received training on the issue that does not appear in the *Enumerator Manual*.

“Observation logs” were used helped to nudge APMs and TPMs to record issues as they observed them in the pilot.

After the pilot, the consultant debriefed the APMs and TPMs from each region. This included a review of the observation logs and a detailed look at the “swap set”, that is, cases whose targeting status differed under wealth ranking versus scoring. This highlighted an indicator that did not make sense to field staff and several cases in which the relevance of an indicator differs between wealth ranking and poverty scoring.

This rest of this section lists issues and discusses their resolution.

### 1. How many members does the household have?

This was the most problematic indicator. Several households in the swap set had low poverty likelihoods because they had only one or two members, yet were ranked D or E by the village. In general, villagers care less than the scorecard about household size. Indeed, staff sometimes saw larger households as *less* poor, if they had more earners.

Several field staff suggested dropping this indicator. I declined, however, because, in the data, it is the single most-powerful predictor of consumption-based poverty. This is not a fluke, as it holds in almost all of the other 40+ countries with poverty scorecards. The strongest feature of the data is that larger households are more likely to be poor.

Nevertheless, this indicator sometimes leads the scorecard to make big mistakes. This underlines the need for a process to manage requests for exceptions.

Most enumerators failed to read the definition of *household member* to the respondent, so this was emphasized in later training.

As usual, there were several questions about the determination of household members:

*Do sons staying at a monastery count?* Sometimes monks are supported by their parents, others are supported by relatives or from donations. According to the IHLCA Technical Committee, monks count as members of their parents' household if they have a definite planned date of return, regardless of how they are supported.

*Do hired workers who eat and sleep with the household count?* It is common for food and shelter to be part (or all) of the pay of hired workers. Because hired workers do not “share” expenses with the household, they do not count as household members.

*Does a family member who works elsewhere count?* According to the IHLCA Technical Committee, migrant workers count as household members if:

- They consider themselves household members, and
- They contribute economically to the household

2. In the last year, how many household members in their main occupation were skilled agricultural and fishery workers or in elementary occupations?

This was the second-most controversial indicator. Although it was not flagged in the indicator review nor in the pre-pilot training, field staff consistently said that more workers—even if in agriculture or elementary occupations—implies less poverty. Villagers seem think in terms of *total household income*, not *per-capita consumption*.

Furthermore, this indicator often stood out for cases in the swap set. At least for households with many people but low dependency ratios, having more workers does seem to be linked with less poverty.

*Do own-account farmers count as skilled agricultural or fishery workers?* Yes.

*What is an elementary occupation?* This is unskilled physical labor, done not as a business but as day labor (sometimes inexplicably called “casual labour”). Most workers in elementary occupations do not furnish their own tools, are hired for a specific task or for a short time, and they do not do the same type of work in all of their engagements.

After the pilot, I dropped this indicator, even though this increased the points assigned to small households. Its relationship with poverty, while perfectly logical in theory and obvious in the data, was the opposite of what staff and villagers expected. I replaced it with a simple, safe (if less powerful) alternative, the type of walls of the home.

3. What is the highest standard/diploma/degree that the female head/spouse has passed?

Field staff did not like this indicator, although they disliked the first two more. Few female head/spouses in the pilot had any education, so this indicator did not help to distinguish among household's poverty levels. For the few female head/spouses who reached second standard or beyond, field staff complained that it was not fair to get a lot of points for education when the female head/spouse still spent almost all her time in basic chores, just like her less-educated neighbors.

*Does it matter why the female head/spouse has no education (for example, her village had no school when she was a girl)?* No.

*The father ran off with another woman, and then the mother died, leaving the oldest teen-age son to care for his two little brothers. Who is the female head/spouse?* The teen-age head is male, and he has no spouse, so there is no female head/spouse.

4. How many rooms does the household occupy, including bedrooms, living rooms, and rooms used for household businesses (do not count toilets, kitchens, balconies, and corridors)?

Field staff accepted this indicator.

*What is a room? Does a simple hut, partitioned into two areas by a curtain or a cabinet, count as one room or two?* A room is an area with a distinct purpose that is separated from another by a partition, be it load-bearing or not. So the example would count as two rooms.

*Does a storage room with sacks of paddy etc. count?* According to the IHLCA Technical Committee, storage rooms are “rooms used for business” and should be counted.

*Does a storage room in a building separate from the main dwelling count?* According to the IHLCA Technical Committee, it still counts.

*Some houses have a lean-to on one or several sides of the main dwelling. Do these count as rooms?* No, a room must have be enclosed by walls on all sides, not just a roof.

*Two households share a dwelling. How should the rooms be counted?* According to the IHLCA Technical Committee, a room can be assigned to only one household, and each room should be assigned to the household that uses it the most.

*What if two households say they use a room equally?* The “Guidelines” and the IHLCA Technical Committee do not say. Enumerators should use their own best judgment.

*If two households say they use a room equally, can it be assigned to the household with the most members?* Yes, if that is your best judgment.

5. What is the major construction material of the floor (observe, do not ask)?

In the pre-pilot ToT, the APMs and TPMs quickly noted that “earth/sand” is the lowest-quality type of floor and so should be option A or joined with others in option A. This makes sense, the points were very close (0 and 1), and dirt floors are not common anyway. So I put “earth/sand” into option A.

Otherwise, this indicator did not cause problems. Field staff did seem to make a bit of a show, pausing and silently looking around pointedly at this stage in the interview. If they cannot look around more discreetly, then—to be polite—they should tell the respondent “I’m just looking at the floor now to see what kind it is”.

*Can I ask the household if I cannot see the whole floor?* Yes.

*Should I determine the major type of floor material by square footage or by the area that the family spends the most time on?* The “Guidelines” and the IHLCA Technical Committee do not say, so use your own best judgment.

*What is the difference between wood planks and finished wood?* Finished wood has been smoothed or painted; wood plants are rough and unpainted. In any case, it does not matter, because they are both part of the same response option.

6. What type of stove is used most often for cooking food in the household?

This indicator did not cause problems.

The original “Guidelines” defined an *improved stove* as one that requires only a single stick. That definition made the APMs laugh; they said that, years ago, the ads for the stoves claimed that they needed only one piece of wood. The APMs noted that firewood varies in size, and some sticks are too small even with 100-percent efficiency. In the field, the APMs made it a point to ask households with improved stoves about the number of sticks required, and the respondents—puzzled by the question—said “four or five”. The “Guidelines” was edited to reflect reality.

7. Does any member of your household own or have access to a wooden food-storage cabinet (including one rented to others or pawned)?

The pilot quickly revealed that while many households did not own a piece of furniture dedicated for use as a food-storage cabinet, many households had cupboards with shelves where they placed food behind a closed door to keep it safe from cats, rats, insects, and small children. According to the IHLCA Technical Committee, cupboards used as food-storage cabinets are to be counted as food storage cabinets.

*Does a plastic dome formed like a solid net, placed over leftover plates of food, count as a food-storage cabinet? No.*

*Does an old, hand-made, bamboo food-storage cabinet count? Yes.* The relevant IHLCA survey questions do not distinguish among assets’ quality, just their presence/absence.

*The household says that it has no food-storage cabinet because it never has any left-over food to store. What response should be marked? “A. No”, because the household does not have a food-storage cabinet. The reason for (not) having it is irrelevant.*

In the pilot, field staff and APMs echoed raters’ earlier claims that food-storage cabinets are weak indicators of poverty. I had resisted this, because it is a small asset related to women’s work and because the data say it is a good predictor. Nevertheless, I relented when the APMs asked to change the indicator to ask about a cupboard and/or a food-storage cabinet. For this new question that appears in the final scorecard, households who own or possess a cupboard but not a dedicated food-storage cabinet are counted as owning both assets if the cupboard is used as a food-storage cabinet.

*What is the definition of a working/functioning food-storage cabinet? If it keeps animals and children away from left-over food, then it works and should be counted.*

8. Does any member of your household own or have access to a black-and-white or colour TV (including one rented to others or pawned)?

In the pilot, no issues surfaced for this indicator.

9. Does any member of your household own or have access to a bicycle, boat, motorcycle, or motorcar (4 wheels) (including one rented to others or pawned)?

This indicator was well-accepted. APMs asked to add other transport assets:

- Carts
- Motorized boats
- Power tillers
- Tractors
- Trawlagees
- Trishaws

These were originally omitted because they are from an IHLCA module on “Business Assets” that I thought applied only to households with non-agricultural businesses. But the IHLCA Technical Committee later made clear that it applied to all households. Thus, I added the requested transport assets. Carts had to be left out because the IHLCA asks only about carts used for non-agricultural businesses, but most carts are used for agricultural and personal transport.

10. What is the total area (in acres) of all plots of land (agricultural, forest, pasture, for livestock breeding, or water surfaces) that any member of the household has the right to use?

This indicate had a number of issues. First, some field staff asked how to mark households without any agricultural land (as “A. None to 1.9”).

Second, it was difficult to interpret “right to use”. This was resolved by defining it as the “right to exclude others from use”. Thus—like the other assets—the concept covers both owning (even if land is rented out, lent out, or sharecropped out) and possessing (even if land is rented in, borrowed in, or sharecropped in). Shifting cultivators have the “right to use” land if—at the time of the survey—they can prevent other households from using it. Households that use land illegally (for example, making charcoal in protected reserves) do not have “right to use”. Households (or their sub-lessees) with

fishing permits for a specific stretch of river can keep others out and so have the “right to use”. In contrast, households who fish in a place where others can also fish—such as the open seas—are not counted as having the “right to use” the “land” because then everyone who used it would be counted as if they controlled the whole thing.

Third, field staff worried about regional variation in the quality of land. For example, one acre of irrigated *le* (paddy) land in the Delta has a different relationship with poverty as one acre of rain-fed land in Shan. APMs understood that on average for Myanmar overall, more land mean less poverty, but from the rating exercise through the end of the pilot, all users consistently expressed concerns about this indicator.

Fourth, field staff wished for more granular response options. They did not like, for example, that 1.8 acres was associated with the same points being landless, nor that someone with 2.1 acres got many more points than someone with 1.8 acres. Scorecard construction started with about 10 narrow ranges, but the statistics suggested that they be combined into a few broad ranges.

I had originally included land (as well as the number of agricultural laborers and workers in elementary occupations) because Myanmar is predominantly agricultural, and the two main assets of most of its poorer people are labor and land. Nevertheless, I dropped this indicator because it consistently troubled users.

To replace it (and to also compensate for dropping the number of agricultural laborers), I sought another agriculture-related indicator. The outcome embeds three questions:

- Are any household members involved in agriculture as their main job?
- If yes, does the agricultural household have any land?
- If yes, does the landed, agricultural household have any large, non-draught farm animals (mainly, in practice, cows)

The first question identifies non-agricultural (usually, non-rural) households. The second question identifies landless agricultural households, the poorest type. The third question identifies landed agricultural households who raise large animals for food.

While complex, the APMs preferred this to the original land indicator. It captures three key aspects of agriculture (presence, land, and livestock), and the points make sense.



The new indicator led also to a number of questions.

*Does garden land count?* Yes.

*Horses, and donkeys/mules are always draught animals. Do they count?* Yes. Draught oxen and draught buffalo do not count, but draught mythun, horses, donkeys, and mules do, regardless of draught status.

*Does a draught cow count?* No.

*What if a household member's main job is gathering forest products?* Such a household is agricultural. *Agricultural* is shorthand that encompasses working in agriculture, hunting, fisheries, forestry, mining, and quarrying. These are all fundamentally extractive sectors, harvesting from the earth, water, and sun/weather.