

附件 II-C: 评估成果和产出绩效的方法

1. 本附件说明了世界粮食计划署用于展示成果和产出绩效结果数据的方法。

1. 按成果评估世界粮食计划署的方案绩效

1.1. 按战略目标分列的成果绩效

2. 2022 年，采用了经修订的《全组织结果框架（CRF）2017 - 2021》中定义的方案成果链方法，在战略目标层面评估了世界粮食计划署的方案绩效，及其对可持续发展目标 2 和 17 的贡献。评估是根据报告年度内活跃行动的总体成果绩效而得出。本分析以国家办事处在 2022 年监测的全组织成果指标为基础。

3. 2022 修订了用于评估成果绩效的方法，以解决与所有指标的绩效应用同等权重有关的局限性，而不论行动规模如何。这些更改会影响下面概述流程中的第 3 步，而第 1 和第 2 步则保持不变。

第 1 步 - 确定分析要考虑的指标

4. 世界粮食计划署对每项方案中监测的每个成果指标、每个目标群体、地点和交付方式进行评级。每个独特的组合都称为“测量”。考虑了以下事项：

- 只包括作为经修订的《全组织果框架》一部分的指标的测量。不包括提供国家一级成果分析（“国别成果指标”）的个别年度国家报告中发现的那些测量。
- 只报告完整的测量 — 那些有基线、年度目标和随访值的测量（见图 1）。如果缺少一个值，则从分析中剔除结果指标测量值。如果报告年度记录了一个以上的随访值，则分析中只包括最后的值（见表 1）。

图 1: 结果指标数据点和组合



表 1: 在特定地点的特定活动下援助的特定目标群体的成果指标测量实例

成果指标	类别	组合	形态	类型	方向	基线	2022年目标值	2022年后续价值
粮食消费计分	可接受粮食消费计分的家庭百分比	Adamawa / CAR 难民食品/ URT1/ URT_GD	粮食	全组织	增加	56.76	>80	66.4

缩写: CAR = C 中非共和国; GD = 一般性分发; URT = 无条件资源交付。

c) 当一个指标涵盖一个以上的次级指标类别时，为了避免重复计算，只考虑其中一个类别。以下是具有多个类别的指标的完整列表，分析中仅考虑了缩进的子指标：

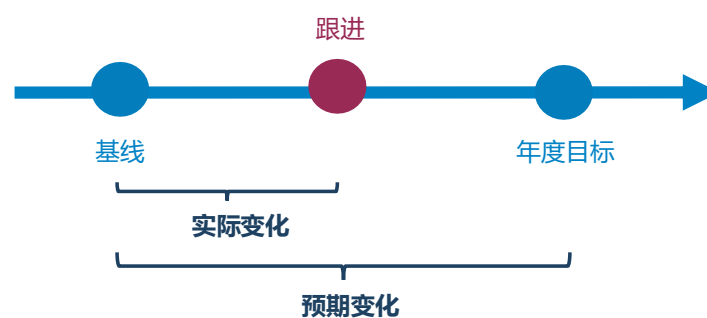
- 粮食消费的分（FCS）：
 - FCS - 具有可接受粮食消费计分的家庭百分比。
- 粮食消费计分 - 营养：
 - 从未食用过富含血红素铁的食物家庭百分比（在过去七天内）。
 - 从未食用过富含蛋白质食物家庭百分比（在过去 7 天内）。
 - 从未食用过富含维生素 A 食物家庭百分比（在过去七天内）。

- 基于生计的应对策略指数：
 - 未使用基于生计的应对策略的家庭百分比。
- 就读率/辍学率：
 - 就读率。
- 通过世界粮食计划署支持的汇集系统小农实现的销售价值和数量：
 - 价值（美元）。
 - 数量（吨）。
- 按交付方式分列的女性、男性或男女双方共同决定食品、现金或凭证使用的家庭比例。
 - 男女共同决定食物或现金使用的家庭百分比。

第 2 步 - 评估每个结果指标衡量的绩效

5. 通过将成果指标的年终值与年度目标和每项《国别战略计划（CSP）》逻辑框架确定的最新基线进行比较来评估成果绩效。
6. 修订后的 CRF 中有三种类型的指标：
 - **增加指标 (>=)**：干预后预计值会增加的指标。
 - **减少 (<=)**：预计干预后值会下降的指标。
 - **具体背景指标**（无法预先确定方向）：国家办事处根据预期的方案成果确定值方向的指标。
7. **检查每项指标的年度目标的一致性**：世界粮食计划署的基于成果的报告，为全组织一级选定的成果指标设定目标，而其他目标则根据具体业务情况在国家一级设定。与基线相比，当目标反映改进时，认为它们是一致的。然而，目标有时并不代表与基线相比的改进，因此被认为不一致。设定一个反映与基线相比情况恶化的目标可能有行动的原因。根据 2022 年的方法，目标不一致的测量结果将被标记为需要审查，如果发现有效，则在分析中予以考虑。
8. **绩效计算 - 成果公式**：计算成果的公式，将部分由于世界粮食计划署干预而观察到的指标值的实际变化与预期变化（代表在全组织或国家办事处一级确定的年度目标）进行比较。有关具有一致性目标的增长指标示例，请参见图 2。

图 2：成果公式的应用示例



9. 用于计算绩效的公式如下所述：
 - a) 如果目标设定始终与基线一致：

$$\frac{\text{实际变化}}{\text{预期变化}} = \frac{(\text{随访} - \text{基线})}{(\text{年度指标} - \text{基线})}$$

- b) 如果目标设定与基线不一致：

$$\frac{\text{实际变化}}{\text{预期变化}} = \frac{(\text{随访} - \text{年度目标})}{(\text{基线} - \text{年度目标})}$$

10. **计算绩效 - 颜色编码**：每个结果指标的绩效评级，根据如下所示的等级以颜色代编码表示。

3	进展强劲，实现或超过目标 指标的实际变化是预期变化的 80%或以上（80%-80+∞%）
2	有些进展 指标的实际变化在预期变化的 50%到 80%之间（50% - 80%）
1	进展不足或没有进展 指标的实际变化小于预期变化的 50%（50-∞% - 50%）

第 3 步 - 确定每项战略成果类别的全组织成果指标的绩效

- 2022 年，对这一步骤进行了修订，以解决第 3 段中提到的局限性，即指标报告的绩效不能反映相关行动的规模。
- 一旦为每项指标测量的绩效评级分配了颜色代码，就根据与之相关的行动的相对大小调整测量值（即乘以一个权重），然后根据 CRF 结果链进行汇总，以确定整个世界粮食计划署的绩效。
- 每项指标测量的调整，由相关行动的规模决定，并根据以下一项进行计算，具体取决于结果指标：每日交付（食品和现金交付）的次数；用于加强能力或服务的支出金额；或受益人的数量。
- 计算行动规模取决于每日交付次数的指标的每次测量的调整（权重）的公式为：

$$\frac{\text{每项完整测量的交付次数}}{\sum \text{该指标所有完整测量的交付次数}}$$

- 例如，如果与多哥行动的个别测量有关的交付占与该指标相关的所有交付的 2%，则该指标在多哥的绩效将乘以 2%；而在也门与同一单个指标相关的行动，与该指标相关的所有交付占 20%时，将获得 20%的权重。
- 然后为每个成果指标代码聚合加权颜色代码。
- 数据可用性：**对于每项指标，绩效分析只包括那些报告了足够数据的国家办事处。“足够数据”被定义为一套完整的衡量标准，就与给定成果指标相关的交付总额而言，至少占国家业务规模的 50%（取决于指标，以交付、支出或受益人来衡量）。
- 计算给定指标的数据充分性的公式，其行动规模基于一个国家的每日交付次数：

$$\frac{\sum \text{该指标在某国的完整测量中的交付次数}}{\sum \text{该指标在某国所有测量中的交付次数}} \geq 50\%$$

- 例如，如果喀麦隆“干预措施的粮食消费计分 — 具有可接受粮食消费的家庭比例”的完整测量值仅占喀麦隆与该指标相关的交付总额的 40%，那么喀麦隆国家办事处的数据就没有包括在粮食消费计分分析中，喀麦隆被归类为没有报告该指标足够数据的国家。

表 2：战略成果类别 1.1 下绩效报告实例

成果 1.1：保持/加强个人和家庭获得充足粮食的机会

成果指标	报告的国家数	报告了足够数据的国家数	国家报告率	测量数	完整测量数	报告率	绩效评级 ¹			与基准相比显示改善或稳定的测量值数	达到或超过目标的测量数
							2022	2021	2020		
1.1.1 粮食消费计分 - 具有可接受的粮食消费计分的家庭百分比	69	58	84%	176	160	91%	●	●	●	80	46
1.1.10 入学率	30	26	87%	33	33	100%	●	●	●	23	19
1.1.11 就读率（新增）	19	15	79%	21	21	100%	●	●	●	16	14
1.1.18 应急能力指数	1	1	100%	1	1	100%	●	●	●	1	0
1.1.19 参加方案（覆盖面）的合格人口比例	23	21	91%	59	54	92%	●	●	●	44	31

- 附件 II-D 的表格展现了汇总结果。表 2 提供了说明计算的摘录。应用了以下定义：
 - 报告国家数：**根据一项或多项目标群体、地点、活动和/或模式的指标进行报告的国家数目。
 - 报告足够数据的国家数：**报告足够数据的国家办事处数（见关于数据可用性的第 17-18 段）。

- c. **国家报告率**：报告足够数据的国家数除以进行报告的国家总数。
- d. **测量数**：针对指标的测量总数。例如，可以测量三个地点的一个活动和每个地点的三个目标群体的粮食消费计分，总共有 9 个测量值。
- e. **完整测量数**：报告足够数据的国家办事处，报告了所有三个数据点（年度随访、年度目标和基线）的测量数。
- f. **报告率**：完整测量的数量除以测量总数。
- g. **与基线相比显示改善或稳定的测量值的数**：与基线相比，年度随访值有所改善或保持稳定的测量数。
- h. **达到或超过目标的测量数**：年度随访值达到或超过年度目标值的测量次数。

1.2. 按方案领域分列的成果绩效

21. 按方案领域评估成果绩效的方法与按战略目标评估成果绩效的方法相同。根据以下指标评估方案领域的绩效。

1.2.1. 针对营养的活动和营养敏感型干预措施

- 22. 为针对营养活动和营养敏感方案编制，评估营养的成果绩效。
- 23. 针对营养的活动包括治疗急性营养不良方案，包括艾滋病毒/肺结核保健和治疗，以及预防营养不良的方案编制，包括预防急性营养不良、发育迟缓和微量营养素缺乏症。
- 24. 营养敏感型方案编制有助于改善营养结果，并补充其他类型的干预措施，如一般性交付、基于学校的方案、粮食系统和小农支持，以及资产创造和生计支持。采用具体指标评估干预类型的营养成果绩效。
- 25. 用于计算绩效表的成果指标值以经修订的 CRF 中列出的营养成果指标为基础，并针对具体的次级活动进行监测。

针对营养的活动

营养治疗指标	子活动
中度急性营养不良治疗恢复率	中度急性营养不良的治疗
中度急性营养不良治疗死亡率	中度急性营养不良的治疗
中度急性营养不良治疗无反应率	中度急性营养不良的治疗
中度急性营养不良治疗违约率	中度急性营养不良的治疗
参与方案的人口比例（覆盖面）	中度急性营养不良的治疗 食疗（治疗严重急性营养不良）
抗逆转录病毒治疗违约率	艾滋病毒/结核病保健和治疗
结核病治疗违约率	艾滋病毒/结核病保健和治疗
抗逆转录病毒治疗营养恢复率	艾滋病毒/结核病保健和治疗
结核病治疗营养恢复率	艾滋病毒/结核病保健和治疗

营养预防指标	子活动
最低饮食多样化 - 女性	预防发育迟缓
接受最低可接受饮食的 6-23 个月龄儿童比例	预防发育迟缓
参与方案的合格人口比例（覆盖面）	预防微量营养素缺乏症 预防急性营养不良 预防发育迟缓 独立的补充微量营养素
参与足够数量分发的目标人口比例（遵从率）	预防微量营养素缺乏症 预防急性营养不良 预防发育迟缓

营养敏感型干预措施

营养敏感型指标	子活动
最低标准 - 女性	除预防发育迟缓外的所有内容
接受最低可接受饮食的 6-23 个月龄儿童的比例	除预防发育迟缓外的所有内容
粮食消费计分 - 营养*	全部
优质和营养丰富的食品产量增加百分比	全部
报告营养作物产量增加的目标小农户百分比	全部

* 这是一个具有多个类别的指标，评估中仅包括与“从未消费过的家庭百分比”相关的值。请参阅上文第 11-19 段；关于第 3 步，包括关于汇总的一般规则的第 16-19 段。

1.2.2. 基于学校的方案

26. 衡量基于学校方案的成果绩效，考虑了现场学校供膳、替代带回家日粮和带回家日粮。

27. 用于计算绩效指标表的结果指标值是基于经修订 CRF 中列出的营养成果指标。

学校供膳指标
就读率（新）
入学率
毕业率（新）
在小学两个年级结束时表现出阅读和理解本年级文本能力的学生百分比（新）
就读率/辍学率（新）*
SABER 学校供膳国家能力（新）

* 这是一个具有多个类别的指标，评估中仅包括与“就读率”相关的值。请参阅上文第 11-19 段；关于第 3 步，包括关于汇总的一般规则的第 16-19 段。

SABER：以获得更好教育成果的系统方法。

1.2.3. 小农支持、生计支持、粮食系统投资和风险管理

28. 计算成果绩效是为了评估资产创造和生计、小农农业和市场支持，以及气候变化适应和风险管理活动所取得的集体成就。

29. 用于计算绩效表的指标的成果指标值，以经修订的 CRF 中的以下成果指标为基础。

指标
目标社区中，报告受益于资产基础增强的人口比例
目标社区中，报告受益于增强生计资产基础的人口比例
目标社区中，报告环境效益的人口比例
粮食消费计分 - 具有可接受粮食消费计分的家庭百分比（活动标记小农农业市场支持/气候变化适应和风险管理/资产建设和生计）
平均降低的应对策略指数 - （标记小农农业市场支持、气候变化适应和风险管理或资产建设和生计的活动）
生计应对策略指标 - （标记小农农业市场支持、气候变化适应和风险管理或资产建设和生计的活动）
世界粮食计划署面向小农采购的合同违约率（百分比）
通过世界粮食计划署支持的农民汇集系统销售的目标小农户百分比
世界粮食计划署从小农汇集系统采购粮食的百分比
小农收获后损失率
通过世界粮食计划署支持的汇集系统小农户销售的价值和数量*
有证据表明管理气候震荡和风险的能力有所提高的目标社区比例

*这是一个具有多个类别的指标，其中与“销售价值”和“销售量”相关的值分别汇总，其绩效指标分别用颜色编码。

1.2.4. 无条件资源交付

30. 计算成果绩效，以评估无条件资源交付的结果。

成果指标具体情况

31. 用于计算绩效表指标绩效值是基于经修订 CRF 中的以下成果指标，并监测了一般性分发。

粮食安全指标	活动标记
粮食消费计分	一般性分发
基于生计的应对策略（不使用应对策略的家庭的比率）	
基于消费的应对策略指数（平均）	
满足基本需求的经济能力（新）	

1.3. 按跨领域结果分列的成果绩效

32. 跨领域结果的成果绩效评估，采用了按战略目标评估成果绩效，以及按方案领域评估绩效相同的方法。

1.4. 成果分析的方法局限性

33. 尽管在 2022 年进行了修订，但该方法仍然存在两个局限性。首先，当多项活动被指派相同的预算编号（对于按支出加权的少数指标）或按交付总额比例加权的指标具有多个类别时，用于估计行动规模的方法具有局限性。然而，这些局限的影响是有限的；在第一种情况下，有关测量只是分析中包括的所有测量的一小部分，在第二种情况下，因为国家办事处记录的测量方式不大可能产生类别之间交付次数的差异。第二个局限性与数据可用性有关：绩效分析只包括那些报告了足够数据并拥有相应行动规模数据的国家办事处。虽然这有助于世界粮食计划署避免从不完整的数据中得出结论，但这意味着绩效评级可能并不反映该组织的整体业务，而可能只反映一部分开展相关业务的国家的绩效。为了缓解这种问题，每项战略目标下的绩效评级报告都提到了分析中包括的国家数量。评级应与随附的叙述性解释一起阅读，以便更好地概述绩效。

2. 在产出层面评估世界粮食计划署的绩效

34. 世界粮食计划署通过其产出指标获取短期方案结果。除了援助的受益人和分发交付的数量外，产出指标还可以衡量创建、恢复或改进的受援助社区的数量、卫生中心、卫生设施、学校菜园等。

35. 目前应用近 400 个产出指标和 60 个测量单位来监测产出。为了评估全组织一级的绩效，将一些使用相同计量单位的产出指标组合在一起。只汇总经修订的 CRF 中包括的产出指标，分析中只考虑实际值和非零计划值的指标。为便于将年度绩效报告重点放在全组织层面的（非具体国家）指标上，只包括完整衡量次数或相关国家大于 1 的指标。

表 3：产出指标测量示例

产出类别 A：交付的资源									
A.5 分发的非食品物质数量	182	137	75%	34	30	88%	8,559,313	6,797,896	79%
A.6 协助的机构站点数目									
援助的机构站点数量 - 厕所	7	4	57%	7	4	57%	1,788	1,378	77%