Identifying "Good" Polygenic Risk Scores: A Dual-Lens Framework

Context: Phenome Health — Kraken Knowledge Graph & ARK v1.0 Return-of-Results **Date:** February 2026

The Problem

Not all polygenic risk scores are created equal. The PGS Catalog contains over 4,000 scores built with widely varying methods, sample sizes, and levels of validation. Phenome Health needs to evaluate PRS quality from two complementary perspectives:

1. **Kraken (Knowledge Graph):** Which PRS models deserve nodes in the graph, and which genes should they link to? The concern here is data modeling fidelity — avoiding noisy or unreliable edges that degrade graph quality.
2. **ARK v1.0 (Return-of-Results):** Which PRS are ready for clinical decision support and patient-facing reports? The concern here is clinical validity, actionability, and responsible communication of genetic risk.

These two lenses overlap substantially but diverge in important ways. This document proposes a unified framework that serves both, with clear annotations on where criteria apply to one context, the other, or both.

Two Perspectives, Shared Foundation

The ARK v1.0 inclusion criteria [7] define three pillars for clinical PRS: validation and discrimination, clinical impact beyond conventional risk factors, and actionability with evidence-based thresholds. The Kraken framework, developed from PRS-RS reporting standards [2] and PGS Catalog metadata [1, 4], focuses on statistical construction quality, ontology mappability, and gene-level decomposition.

The following table summarizes where they converge and diverge:

Concern	Kraken (Graph)	ARK (Clinical)	Overlap
LD-aware construction	Required	Implicit (large, recent cohorts)	Partial
External validation	Required	Required (replication across cohorts)	Full
Ancestry appropriateness	Strongly preferred	Required (diverse cohorts, continuous ancestry)	Full
Risk gradient / discrimination	Strongly preferred (AUC, OR)	Required (top 5–10% elevated risk)	Full
Ontology mapping (EFO/MONDO)	Required	Not addressed	Kraken only
Scoring file availability	Required	Not addressed (may build in-house)	Kraken only
Clinical impact beyond risk factors	Not addressed	Required	ARK only
Actionability / clinical thresholds	Not addressed	Required	ARK only
Interpretability / patient-facing	Not addressed	Required (with genetic counseling)	ARK only
Regulatory compliance (FDA/LDT)	Not addressed	Required consideration	ARK only
Gene-level decomposition	Required (for graph edges)	Not addressed	Kraken only

The key insight: every PRS that passes ARK's clinical bar should also pass Kraken's quality bar, but not vice versa. Kraken can include exploratory or research-grade scores that aren't yet ready for clinical return-of-results.

Background: Why LD-Awareness Matters

A PRS is a weighted sum of allele dosages multiplied by effect sizes estimated from GWAS. The challenge is that nearby variants are correlated due to linkage disequilibrium — if you naively sum their effects, you double-count the same causal signal and inflate the score.

LD-aware methods address this by shrinking or re-weighting effect sizes to account for correlation structure. The PGS Catalog documents over 50 unique construction methods [1, 4]. The most common LD-aware

approaches include:

Method	Approach	LD Handling Quality
PRS-CS / PRS-CSx	Bayesian continuous shrinkage	High
LDpred2	Bayesian regression	High
lassosum	Penalized regression	High
SBayesR	Bayesian mixture model	High
LDpred-funct	Functional annotation + Bayesian	High
metaGRS	Meta-scoring across multiple PRS	High
C+T / P+T	Clumping + Thresholding	Moderate
PRSice2	Pruning + Thresholding	Moderate

Scores built with clumping-and-thresholding alone are weaker candidates unless they have been extensively validated, because they discard variants rather than properly modeling their joint effects. Bayesian methods (PRS-CS, LDpred2) generally produce better-calibrated scores and are preferred for clinical applications [4, 5].

The ARK criteria do not explicitly name LD-aware methods, but the requirement for scores "created using large, recent, diverse cohorts" [7] effectively selects for modern construction methods, since recent high-quality PRS publications overwhelmingly use Bayesian or penalized regression approaches.

Background: What Validation Means

External validation means testing a PRS in a cohort that was not used to develop or train the score. The ClinGen PRS Reporting Standards (PRS-RS) define a 22-item framework for evaluating PRS quality, with particular emphasis on reporting discrimination, predictive ability, and calibration in independent validation samples [2].

Both Kraken and ARK require external validation, but with different emphases:

Kraken's concern is that unvalidated scores may produce unreliable gene-level edges. If a PRS has never been tested outside its training data, the variant weights — and therefore the gene contribution scores we derive — may be overfit.

ARK's concern is clinical: the score must demonstrate "a clear gradient of increasing risk, with an upper stratum (e.g. top 5% or 10%) exhibiting materially elevated relative or absolute risk compared to the population mean" [7]. This is a stronger requirement than simply having an independent evaluation — it demands that the score meaningfully stratifies patients.

Shared validation concerns include:

- **Ancestry transferability:** Scores developed in European-ancestry GWAS often show reduced predictive accuracy in other populations due to differences in LD structure, allele frequencies, and genetic architecture [1, 5]. PGS Catalog benchmarking in UK Biobank confirmed that many European-derived scores were not significantly predictive in South Asian or African ancestry participants [1]. The ARK criteria go further, specifying that ancestry should be treated as a continuous spectrum with PRS distribution adjustments (mean and standard deviation), ideally leveraging WGS from diverse cohorts like All of Us [7].
- **Reproducibility:** The PGS Catalog Calculator (pgsc_calc) provides a standardized pipeline for computing scores with ancestry estimation and normalization, enabling reproducible application across cohorts [4].

Unified Quality Criteria

The following criteria merge the Kraken and ARK perspectives into a single framework. Each criterion is tagged with its applicability.

Tier 1: Required for Kraken Node Inclusion

These four criteria must all be met for a PRS to receive a node in Kraken. They represent the minimum bar for data quality in the knowledge graph.

1. LD-aware construction method [Kraken] [ARK: implicit] The score must use a method that accounts for linkage disequilibrium when assigning variant weights. The PGS Catalog method_name field documents the algorithm used for each score [1, 4].

```
Filter: method NOT IN ['lead SNPs only', 'simple sum']
Preferred: method IN ['PRS-CS', 'LDpred2', 'lassosum', 'SBayesR', 'metaGRS']
```

2. External validation in an independent cohort [Kraken] [ARK] At least one evaluation must exist in a sample independent from the discovery GWAS. PRS-RS requires reporting of discrimination and calibration metrics in validation samples [2, 6].

```
Filter: n_evaluations >= 1 WHERE evaluation.independent = true
```

3. Scoring file available with harmonized coordinates [Kraken] Variant-level information (rsID or chr:pos, effect allele, weight) must be available in GRCh37 or GRCh38 coordinates. This is required for computing gene mappings via chromosome position — the foundation of our PRS → Gene edges. Note: ARK may develop scores in-house on a case-by-case basis [7], in which case scoring files would be generated internally.

```
Filter: scoring_file EXISTS AND genome_build IN ['GRCh37', 'GRCh38']
```

4. Mapped to EFO/MONDO trait ontology [Kraken] The predicted trait must map to a structured ontology term for integration with Kraken's existing disease and phenotype nodes. PGS Catalog maps each score to Experimental Factor Ontology (EFO) terms [1, 4].

```
Filter: trait_efo IS NOT NULL
```

Tier 2: Strongly Preferred for Kraken / Required for ARK

Meeting these criteria elevates a PRS from research-grade to clinically relevant. For ARK return-of-results, these are effectively required.

5. Ancestry-appropriate validation [Kraken: preferred] [ARK: required] Validation should exist for ancestry groups relevant to Phenome Health's cohorts (Arivale, UK Biobank, Israeli10K). PRS-trait associations are deflated when the target population differs from the development population in allele frequencies or LD structure [4, 5]. ARK further requires that ancestry be treated as a continuous spectrum with distribution adjustments, ideally using WGS from a large diverse reference like All of Us [7].

```
Filter: ancestry_validated INTERSECTS target_cohort_ancestry
ARK additional: continuous ancestry adjustment implemented
```

6. Multiple independent evaluations [Kraken: preferred] [ARK: required] Scores replicated across multiple cohorts, ideally across different ancestries, provide stronger evidence of generalizability [2, 6, 7].

```
Filter: n_evaluations >= 2, ideally across different ancestries
```

7. Clinically meaningful discrimination [Kraken: preferred] [ARK: required] For binary traits, we prefer AUC > 0.6 or an odds ratio > 2.0 comparing top vs. bottom decile. For quantitative traits, incremental $R^2 > 1\%$ beyond covariates. ARK specifically requires "a clear gradient of increasing risk, with an upper stratum exhibiting materially elevated relative or absolute risk" [7]. Weak scores add noise to the graph without meaningful clinical signal [2, 5].

```
Filter: AUC > 0.6 OR OR_top_decile > 2.0 OR delta_R2 > 0.01
ARK additional: demonstrate clear risk gradient with elevated upper stratum
```

Tier 3: ARK Clinical Requirements (Beyond Kraken)

These criteria are specific to clinical return-of-results and do not apply to Kraken node inclusion. However, Kraken can store these as node properties to support downstream ARK queries.

8. Clinical impact beyond conventional risk factors [ARK] The PRS must add predictive value that meaningfully refines risk stratification beyond conventional clinical risk factors (age, sex, family history, biomarkers, etc.) [7]. This is assessed by comparing model performance with and without the PRS in a multivariable risk model.

> ARK filter: incremental AUC or NRI significant when added to clinical model
> Kraken property: clinical_impact_demonstrated (boolean)

9. Actionability with evidence-based thresholds [ARK] There must be clear actionable thresholds that, when used alongside demographic and clinical risk factors, inform evidence-based decision-making and interventions. Ideally, actionable recommendations exist from professional groups or national expert guidelines [7].

> ARK filter: actionable_thresholds_defined = true
> ARK filter: guideline_recommendations EXIST
> Kraken property: has_clinical_guidelines (boolean)

10. Interpretability and patient-facing readiness [ARK] Results must be translatable from raw scores into clinician- or patient-facing insights. Support systems like genetic counseling should be available to navigate complex risk data and avoid genetic determinism [7].

> ARK filter: interpretability_framework EXISTS
> Kraken property: ark_approved (boolean, set by clinical team)

Tier 4: Enrichment Metadata

These are not filtering criteria but additional properties that enhance PRS node value in Kraken.

11. Gene contribution variance [Kraken] Pre-computed or computable gene-level importance scores using the XPRS methodology (variance of gene contribution score across a reference population). This determines which PRS→Gene edges to create in Kraken. The variance of CSgene is equivalent to per-gene SNP heritability and can be derived from the scoring file plus a reference genotype panel (e.g., 1000 Genomes ancestry-matched subset) [3].

12. GWAS Catalog association linkage [Kraken] Linking the PRS back to its source GWAS study enables provenance edges (PRS → GWAS Study → Lead Variants) and connects to existing Kraken GWAS infrastructure [1, 2].

13. pgsc_calc compatibility [Kraken] [ARK] Scores that can be calculated using the PGS Catalog's standardized calculator pipeline ensure reproducibility and enable batch computation across Phenome Health cohorts [4]. Also relevant for ARK's technical and analytical validity requirement [7].

Computed Quality Tiers

Rather than making a hard quality cutoff at ingestion, we propose annotating each PRS node with a computed quality tier. This allows Kraken to serve both research and clinical use cases from the same graph.

Tier	Criteria	Kraken Use	ARK Eligible?
Gold	All required + ≥2 preferred + ARK criteria met	Full graph integration	Yes
Silver	All required + ≥2 preferred	Research queries, cohort stratification	Candidate (needs ARK review)
Bronze	All required + ≥1 preferred	Exploratory analysis	No
Copper	All required criteria met	Hypothesis generation only	No
Excluded	Missing any required criterion	Not ingested	No

Gold-tier scores are those that pass both the Kraken statistical quality bar and the ARK clinical readiness bar. The clinical team can promote Silver-tier scores to Gold after manual review of clinical impact, actionability, and interpretability.

ARK Implementation Considerations

The ARK v1.0 document [7] identifies four implementation considerations that sit outside the scoring criteria but affect how PRS are deployed. These are captured here for completeness and to note where Kraken can support them.

Technical & Analytical Validity: Genetic measurements must be consistent across platforms. ARK notes that decisions on using pre-existing scores (e.g., from PGS Catalog) versus in-house development will be made case-by-case [7]. Kraken's scoring file requirement (criterion 3) ensures that any PGS Catalog score we ingest has the variant-level detail needed for either approach.

Bias Mitigation (Ancestry): ARK requires ancestry as a continuous spectrum with distribution adjustments [7]. This aligns with pgsc_calc's Z-normalization methods [4], which Kraken can flag via the `pgsc_calc_compatible` property. The recommendation to use WGS from All of Us as a diverse reference panel is relevant for both Kraken's gene contribution variance computation and ARK's clinical score normalization.

Interpretability and Ethics: ARK requires translation of raw scores into patient-facing insights with genetic counseling support [7]. While this is outside Kraken's scope, the XPRS gene-level decomposition [3] stored in Kraken's INVOLVES_GENE edges could directly support interpretability — clinicians could explain "your

elevated T2D PRS is primarily driven by variation in KCNQ1 and CDKAL1" rather than presenting an opaque number.

Regulatory Infrastructure: The FDA regulatory landscape for clinical PRS remains unsettled. The 2024 final rule intended to extend medical device requirements to LDTs was vacated by a federal court in 2025, leaving most PRS tests under CLIA oversight focused on analytical validity [7]. GENinCode's De Novo classification request for a coronary artery disease PRS (CIC-Score) could establish a regulatory precedent if successful [7]. Kraken can store a `regulatory_status` property on PRS nodes to track this evolving landscape.

Proposed Node and Edge Schema

PRS Node Properties

```
// --- Identity ---
pgs_id:          "PGS000013"          // PGS Catalog identifier
name:            "PRS for T2D"        // Human-readable name

// --- Trait Mapping ---
reported_trait:  "Type 2 Diabetes"    // Reported trait string
trait_efo:       "EFO_0001360"        // EFO mapping → Disease node
trait_mondo:     "MONDO:0005148"      // MONDO mapping → Disease node

// --- Construction Quality ---
method:          "PRS-CS"             // Construction algorithm
ld_aware:        true                 // Derived from method
n_variants:      1103075              // Variants in scoring file
genome_build:    "GRCh38"             // Coordinate system

// --- Validation Evidence ---
n_evaluations:   4                    // Independent evaluations
best_auc:        0.63                 // Best reported AUC (binary)
best_r2:         null                 // Best reported R² (quantitative)
or_top_decile:   2.4                  // Top vs bottom decile OR
ancestry_dev:    ["EUR"]              // Development ancestry
ancestry_val:    ["EUR", "EAS"]       // Validated ancestries
has_ext_valid:   true                 // External validation exists

// --- Provenance ---
publication:     "PMID:38302567"      // Source publication
source_gwas:     "GCST90012345"       // Source GWAS study
pgsc_calc_ok:    true                 // Calculator compatible
```

```
// --- Quality & Clinical Readiness ---
quality_tier:      "gold"              // Computed: gold|silver|bronze|copper
ark_approved:      true                // Clinical team sign-off
clinical_impact:   true                // Adds value beyond risk factors
has_guidelines:    true                // Professional guideline support
regulatory_status: "LDT/CLIA"          // Current regulatory pathway
```

Edge Types

```
PRS --[PREDICTS]--> Disease/Trait        (via EFO/MONDO)
  Properties: evidence, best_auc

PRS --[INVOLVES_GENE]--> Gene            (top N by contribution variance)
  Properties: contribution_variance, rank, n_variants_mapped

PRS --[DERIVED_FROM]--> GWAS_Study       (provenance)
  Properties: role

Gene --[existing edges]--> Pathway, Protein, Variant, ...
```

The (INVOLVES_GENE) edge uses the XPRS gene contribution score variance [3] to select which genes receive edges. Only genes above a variance threshold (or top N ranked) are included, avoiding the approximately 3 billion variant-level edges that would result from linking each PRS to all its constituent variants.

Implementation Notes

Data source: PGS Catalog bulk download (FTP) provides scoring files with harmonized coordinates and metadata CSV files with method, evaluation, and trait information. The API can be used for incremental updates.

Gene mapping: Chromosome positions from scoring files are mapped to genes using positional proximity (default 200kb window), supplemented by combining SNP-to-gene (cS2G) mapping that integrates eQTL and Hi-C interaction data [3].

Reference population: 1000 Genomes Project ancestry-matched subsets can serve as the reference for computing gene contribution score variance without requiring individual-level data from Phenome Health cohorts. For ARK applications, All of Us WGS data could serve as a more diverse reference panel [7].

Scale estimate: Based on current PGS Catalog contents, applying the required criteria would likely yield several hundred PRS nodes. With a top-50-genes-per-PRS edge policy, this produces on the order of 10,000–50,000 INVOLVES_GENE edges — well within Kraken's capacity.

ARK promotion workflow: New PRS enter Kraken at Bronze or Silver tier based on automated quality checks. The clinical team reviews Silver-tier candidates against ARK criteria 8–10 (clinical impact, actionability, interpretability) and promotes qualifying scores to Gold with `ark_approved = true`.

References

[1] Lambert SA, Gil L, Jupp S, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics*. 2021;53(4):420-425. doi:10.1038/s41588-021-00783-5

[2] Wand H, Lambert SA, Tamber C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature*. 2021;591:211-219. doi:10.1038/s41586-021-03243-6

[3] Kim NY, Lee S. XPRS: a tool for interpretable and explainable polygenic risk score. *Bioinformatics*. 2025;41(4):btaf143. doi:10.1093/bioinformatics/btaf143

[4] Lambert SA, de Haan HG, Ardlie KG, et al. The Polygenic Score Catalog: new functionality and tools to enable FAIR research. *Nucleic Acids Research*. 2024;52(D1):D1239-D1245. doi:10.1093/nar/gkae528

[5] Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*. 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1

[6] Kuo CL, Lennon R, Engert JC, et al. Implementing Reporting Standards for Polygenic Risk Scores for Atherosclerotic Cardiovascular Disease. *Current Atherosclerosis Reports*. 2023;25:583-592. doi:10.1007/s11883-023-01135-y

[7] Phenome Health. Inclusion Criteria for Polygenic Risk Scores (PRS) — ARK v1.0 Return-of-Results. Internal document, 2025.