Teaching Genealogists AI™
Instructor: Sylvia Vargas

AI Tools for Research and Organization

Module 2 – Record Matching: Solving Common-Name Problems

Learning Objectives

By the end of this module, you will:

- Understand why common Puerto Rican names create record-matching challenges
- Know how to write a comparison prompt that produces a structured AI analysis
- Recognize the types of evidence AI can identify – and where its evaluation must be verified
- Apply the GPS Three-Layer Framework to evaluate any AI-assisted match

---

The Common-Name Problem in Puerto Rican Genealogy

Puerto Rican genealogy regularly presents three name-matching challenges:

Repeated names across generations. Religious names (José, María, Juan, Carmen) were extremely common, and many families reused the same given name in consecutive generations. A search for "José Rivera" in Utuado between 1880 and 1920 may return dozens of results from the same family.

Name variations across record systems. The same person may appear as *José* in a church record, *Joseph* in a U.S. federal census, and *Joe* on a draft registration card. Surname spelling also shifts: *Jiménez* becomes *Gimenez* or *Jimenez* depending on who wrote the record and when.

Dropped or added surnames. In Puerto Rican naming conventions, a person carries two surnames: the father's first surname and the mother's first surname. Records – especially U.S.-era documents – frequently drop the maternal surname. *Rivera Torres* becomes *Rivera* – or sometimes just *Torres* if the enumerator recorded the wrong surname first.

AI handles all three challenges reasonably well when you give it the right prompt.

How AI Approaches Record Matching

AI does not search databases. It compares the text you provide. When you paste two record transcriptions and ask for a comparison, the AI looks for:

- **Age consistency:** Does the age in Record A match the expected age in Record B given the time elapsed?
- **Geographic consistency:** Are the locations the same municipality, neighboring towns, or a plausible migration route?
- **Family group overlap:** Do the same family members (spouse, children, parents, siblings) appear in both records?
- **Phonetic equivalence:** Are the name spellings phonetically identical or similar enough to be the same name?

- **Surname logic:** Does the apparent name difference follow a known Puerto Rican naming pattern (dropped maternal surname, U.S.-era anglicization)?

AI gives you an analysis, not a verdict. The decision – whether two records belong to the same person – remains yours.

The Comparison Prompt

Use this prompt structure when asking AI to compare two records. Replace bracketed text with your actual transcriptions.

Prompt:

> I am a genealogist researching Puerto Rican families. I have two records that may refer to the same person. Please compare them and evaluate the likelihood that they belong to the same individual. Analyze: age consistency (accounting for years between records), geographic consistency, family group overlap (shared family members), name spelling variations, and Puerto Rican surname conventions (dropped maternal surname, anglicization). List the supporting evidence, the contradicting evidence, and the unresolved questions. Do not reach a conclusion for me – present the evidence so I can evaluate it.
>
> **Record A:** [Paste first record transcription]
>
> **Record B:** [Paste second record transcription]

Worked Example: Census to Death Certificate

Record A – 1910 U.S. Federal Census, Utuado, Puerto Rico:

> José Rivera, head of household, age 25. Wife: María. Son: Pablo, age 3.

Record B – 1945 Death Certificate, Utuado, Puerto Rico:

> José Rivera Torres, age 61. Cause of death: heart failure. Spouse: María Delgado. Parents: Luis Rivera and Elena Torres.

The prompt (using the template above) produces this type of analysis:

Supporting evidence: - Age is consistent: age 25 in 1910 would be approximately 60–61 in 1945. (Match.) - Location is consistent: both records place this person in Utuado. - Spouse name is consistent: "María" in 1910 matches "María Delgado" in 1945. (The 1910 census often recorded only given names for spouses.) - Surname logic: "Rivera Torres" in 1945 is the full double surname. "Rivera" in 1910 is the paternal surname only – consistent with how U.S. enumerators recorded Puerto Rican names.

Contradicting evidence: - No contradictions identified in the information provided.

Unresolved questions: - The 1910 census does not record José's parents. The death certificate names Luis Rivera and Elena Torres as parents. These cannot be confirmed from these two records alone. - Pablo, age 3 in 1910, does not appear in the death certificate. His presence or absence in later records is unknown.

GPS Checkpoint: Evaluating This Match

The AI analysis is useful evidence, but it is not a GPS-compliant conclusion. Before treating these two records as belonging to the same person, apply the Three-Layer Framework:

Source analysis: - Record A (1910 U.S. Federal Census): Derivative source (digitized image of an original enumeration form). Information quality for age and location: primary if José provided it himself, secondary if a neighbor or family member reported for the household. Evidence: direct for name and location, indirect for identity match. - Record B (1945 Death Certificate): Derivative source (digitized image of an original civil record). Information quality for age and location at death: secondary (informant is typically a family member, not the deceased). Information about parents' names: potentially primary if the informant knew the parents directly, secondary if based on family knowledge.

What AI cannot do: AI did not examine the original images. It did not assess handwriting, check for alterations, or note whether the enumerator recorded other nearby households that might add context. It had no information about who the informant was on the death certificate.

Your next step: Verify both records against original images on FamilySearch. Search for Pablo Rivera (born ca. 1907, Utuado) in later censuses to corroborate the family unit. Search for a marriage record for José Rivera and María Delgado in the Utuado civil registry to add a third corroborating source.

Name Variation Quick Reference

Variation Type	Example	Likely Explanation	Action
Maternal surname dropped	Rivera Torres vs. Rivera	U.S. enumerator convention	Probable match – verify with a third record
Phonetic spelling	Jiménez vs. Gimenez	Enumerator transcribed by ear	Probable match – verify age and location
Anglicization	José vs. Joseph	U.S.-era records	Probable match – verify

Variation Type	Example	Likely Explanation	Action
			family group
Given name abbreviation	María de los Ángeles vs. María	Common practice	Probable match – verify family group
Surnames reversed	Rivera Torres vs. Torres Rivera	Enumerator error	Possible match – verify parents
Completely different name	José vs. Juan	Possible error or different person	Investigate further before concluding match

When a variation is not explained by any known pattern, treat it as contradicting evidence and document it explicitly before reaching a conclusion.

When AI Gets the Match Wrong

Sometimes AI will identify a plausible match that is actually two different people, or miss a match because of an unexpected spelling. Watch for these warning signs:

- AI expresses high confidence based only on age and location, with no family group overlap
- The family group in one record (spouse, children) does not appear at all in the other
- Geographic locations are in the same general region but different municipalities with no documented migration route
- The age difference is more than five years with no explanation offered

When you see these signs, treat the match as unconfirmed. Document what you found, record why you cannot confirm it, and search for a third source before reaching a conclusion.

What's Next

In Module 3 – Organizing Your Data with AI, you will learn how to convert raw transcriptions, notes, and narrative text into structured tables, standardized dates, and GEDCOM-ready data using targeted AI prompts.