MongoDB Data Model & Schema Proposal

1. Appropriate Fields & Schema

- patient_id: Unique identifier (String or Integer)
- first_name, last_name: Patient names (String)
- age: Patient age (Integer)
- gender: Patient gender (String)
- date_of_admission: Admission date (Date)
- discharge_date: Discharge date (Date)
- medical_condition: Primary diagnosis (String)
- medications: List of medications (Array of Strings)
- address (optional): Nested object for address details
- contact_info (optional): Nested object for phone/email

2. Embedding vs. Referencing

• Embedding is preferred for simple, related data:
  - Store all patient information in one document.
  - Use an array of medication names (Strings).
• Referencing is used when:
  - You have complex sub-documents (e.g., medication metadata).
  - Multiple patients share the same referenced documents.
• For this CSV-based scenario, embedding everything in one ‘patients’ collection is simpler and more performant.

3. Example Document

```
{
  "_id": ObjectId("609e2f4b8d1f4a3b1023a9bc"),
  "patient_id": "P12345",
  "first_name": "Alice",
  "last_name": "Johnson",
  "age": 67,
  "gender": "Female",
  "date_of_admission": ISODate("2023-02-15T00:00:00Z"),
  "discharge_date": ISODate("2023-02-22T00:00:00Z"),
  "medical_condition": "Hypertension",
  "medications": [
    "Lisinopril",
    "Lipitor"
  ],
  "address": { "street": "123 Maple St", "city": "Springfield", "state": "IL", "zip": "62704" },
  "contact_info": { "phone": "555-1234", "email": "alice.johnson@example.com" }
}
```