**### 1. Pregnancies /ˈpreɡ.nən.si/ (số lần mang thai)/GDM (Gestational /dʒes/ Diabetes /ˌdaɪ.əˈbiː.t̬iːz/ Mellitus /məˈlaɪ.t̬əs/)**

**- Đái tháo đường thai kỳ (cuối tam cá nguyệt 2 ~ tuần 24-28)**

**https://pmc.ncbi.nlm.nih.gov/articles/PMC2827530/?utm\_source=chatgpt.com**

**- Thường sẽ hết sau khi sinh nhưng mẹ có nguy cơ cao bị T2D(Type 2 Diabetes Mellitus) về sau; thai nhi dễ gặp biến chứng (thai to, hạ đường huyết sau sinh...)**

**https://pmc.ncbi.nlm.nih.gov/articles/PMC9016728/?utm\_source=chatgpt.com**

**- Ảnh hưởng đến Age, BMI:**

**- Số lần mang thai nhiều -> Tuổi, tăng cân cao**

**https://obgyn.onlinelibrary.wiley.com/doi/10.1111/aogs.13321?utm\_source=chatgpt.com**

**=> Nguy cơ gián tiếp, quan sát được -> Cần phụ thuộc với các yếu tố cao khác.**

**“Đây là khuyến cáo IADPSG 2010 in trên Diabetes Care. Nhóm tác giả dùng dữ liệu HAPO (~25 nghìn thai phụ làm OGTT 75 g) và chứng minh đường huyết mẹ càng cao thì nguy cơ bất lợi cho mẹ–bé càng tăng theo kiểu liên tục; từ đó họ đề xuất ngưỡng chẩn đoán GDM được nhiều tổ chức (trong đó có WHO) xem xét/áp dụng.”**

**### 2. Glucose (đường huyết 2h, mg/dL)**

**- Quy trình nghiệm pháp dung nạp đường uống (OGTT):**

**- Nhịn đói 8 giờ -> lấy mẫu lúc đói (FPG)**

**- Uống 75g glucose hòa với nước (WHO)**

**- Ngồi nghỉ, không ăn/uống/vận động**

**- Lấy mẫu sau 2 giờ**

**Bảng so sánh người mang thai và không:**

**### 3. BloodPressure (huyết áp tâm trương, mmHg)**

**- Về lâm sàng: Phân loại tăng huyết áp chuẩn cần cả SBP (tâm thu) và DBP (tâm trương). Bộ dữ liệu chỉ có DBP, nên không đủ để phân loại tình trạng tăng huyết áp (cả ở người mang thai lẫn không mang thai).**

**- Về mô hình trên dataset Pima: DBP cho tín hiệu yếu với đích Outcome(≈ 0.065)**

**→ ảnh hưởng không cao.**

**- < 60 mmHG: \*\*Thấp\*\* Có biểu hiện chóng mặt, mệt mỏi do máu không cung cấp đủ oxy.**

**- 60 - 80 mmHG: \*\*Lý tưởng\*\* Tim mạch OK.**

**- 80 - 90 mmHG: \*\*Tiền tăng\*\*. Yếu tố nguy cơ phát triển tăng huyết áp.**

**- >= 90 mmHG: \*\*Tăng\*\*. Cần bác sĩ**

**4. SkinThickness – bề dày da tam đầu (mm)**

Proxy cho mỡ dưới da tại vị trí tam đầu -> liên quan đến đề kháng Insulin

**Vấn đề:** Xuất hiện nhiều values 0 có thể là "missing trá hình" hoặc missing data:

   - Khó/không đo được ở lần khám (đo nếp gấp da cần kẹp caliper đúng vị trí; người đo chưa thành thạo; bệnh nhân không hợp tác…).

    - Một số phòng xét nghiệm (ví dụ Insulin) không thực hiện quy trình này

**Bản chất:** các số 0 này là “giá trị thiếu được mã hoá thành 0” (disguised /dɪsˈɡaɪzd/ missing data) chứ không phải đo lường hợp lý về sinh học.

**Minh chứng trích từ Pearson (2006, SIGKDD Explorations):**

[*https://www.kdd.org/exploration\_files/12-Pearson.pdf*](https://www.kdd.org/exploration_files/12-Pearson.pdf)

Tác giả chỉ ra rằng dù metadata nói “không có missing”, nhưng 5 biến (trong đó có TSF – triceps skinfold thickness) chứa “giá trị 0 không khả thi về sinh học”, nên phải coi là missing*. (Page 3)*

Pearson tóm lược: triceps skinfold (TSF) “~29.6% missing”, còn insulin ~48.7% (con số missing này phản ánh thực tế nhiều dòng ghi 0 cho TSF/INS). *(Page 3)*

A graph of a number of dots

AI-generated content may be incorrect.

**Hình 7**: các số 0 lặp lại tạo một “dải” ở đáy nhưng không bị coi là outlier theo thuật toán thông thường.

A graph of a line

AI-generated content may be incorrect.

**Hình 8:** “đuôi dưới phẳng” chính là dấu hiệu zero “giả”