



**To:** HCHS/SOL Investigators and Analysts  
**From:** HCHS/SOL Diabetes/Metabolic Syndrome Scientific Interest Group (SIG) and HCHS/SOL Coordinating Center  
**Date:** April 14, 2021  
**Re:** HCHS/SOL Diabetes definitions recommendation

The HCHS/SOL Coordinating Center released in July 2020 an update to the Visit 2 Derived Variables Dataset (PART\_DERV\_V2\_INV3). Among other updates, it includes new derived variables for the analyses of diabetes incidence. **In this MEMO we summarize the diabetes definitions available in HCHS/SOL and the recommendation from the Diabetes SIG on which definition to use. For consistency across manuscripts, it is preferable to use the same definition.** The Diabetes SIG recognizes that all definitions will have some misclassification associated with them and that the choice to use a particular definition in a manuscript depends on the manuscript's scientific question.

Based on the information that was collected during the HCHS/SOL baseline visit, four definitions for diabetes have been derived and numbered as definitions 2 to 5 in the order of creation; see details of the definitions in the HCHS/SOL Baseline Derived Variable Dictionary. Briefly,

- (a) **Definition 2 (DIABETES2):** ADA lab criteria (FPG, OGTT and A1c measures) or **diabetes medication use based on scanned medication**. Baseline prevalence paper by Dr. Schneiderman et al. ([Diabetes Care, 2014](#)) uses this definition. However, it is not available at Visit 2 because scanned medications are not coded yet.
- (b) **Definition 3 (DIABETES3):** ADA lab criteria or **self-reported diagnosis**.
- (c) **Definition 4 (DIABETES4):** ADA lab criteria or **self-reported medication use in the last 4 weeks**. Note that medication use refers to the past 4 weeks prior to the clinic visit. It is not intended to obtain the medication use history, instead it is to complement the lab measures.
- (d) **Definition 5 (DIABETES5):** ADA lab criteria or **self-reported medication use in the last 4 weeks, or self-reported diagnosis**.

Ideally, to define diabetes incidence we would use the same algorithm as the one used at baseline. However, questions were asked slightly differently that prevent using the same algorithm directly. For example,

- Both at baseline and at clinic visit 2, **self-reported diagnosis** was asked in the Medical History Form (MHE). However, the question refers to a different time-period. At baseline, the question was: "MHE16. Has a doctor ever said that you have diabetes (high sugar in blood or urine)?" In contrast, at clinic visit 2 the question was: "MHE14. Since our last telephone interview with you, has a doctor or health professional told you that you had diabetes or high sugar in the blood?". Therefore, to capture the self-reported diagnosis at visit 2, we need to also include data from all previous annual follow-up calls when the same question was asked in OPE7 of the Out-Patient Self-Reported Conditions Form.
- At baseline, if the participant confirmed that a doctor has ever said s/he had diabetes then women were asked whether that was during pregnancy only. However, at annual follow-up calls and visit 2 women were not asked the follow up question regarding gestational diabetes. Instead, at visit 2 women provided detailed information of all pregnancies between baseline and visit 2 and whether at any of these they had gestational diabetes.



**The Coordinating Center derived indicator variables to identify diabetes incident cases by visit 2 using definitions 3, 4 and 5, and time to the diabetes incident case for definitions 3 and 5. Please refer to the V2 Analyses Methods document for the proper analytic method depending on the definition.**

The main difference for the various definitions is whether to include information on self-reported diagnosis or self-reported medication use as part of the definition. **In general, the Diabetes SIG and the Coordinating Center recommend definition 5 but recognizes exceptions due to the research question of interest.** If definition 5 is not used, it may be informative to do a sensitivity analyses using it.

**Reasons for RECOMMENDING Definition 5** (ADA lab criteria or self-reported medication use, or self-reported diagnosis):

- The wording of the question on medication use (MUE26c) reduces the concern that the medication was prescribed for an indication other than diabetes (e.g., metformin for PCOS or prediabetes, GLP-1 agonist for obesity, SGLT-2 inhibitor for cardiovascular disease and in near future probably CKD). Specifically, MUE26c states: “Were any of the medications you took during the last four weeks for high blood sugar or diabetes?”
- Self-reported medication use complements the lab measures because these can be influenced by the medication.
- Including self-reported diabetes diagnosis results in a higher number of diabetes cases.
- Primary manuscript on incident diabetes ([MS 348](#) under peer review) lead by Drs. Schneiderman and Cordero uses this definition because it is the most comprehensive and authors wanted to capture all who have ever received a diabetes diagnosis.
- Self-reported diabetes diagnosis is in general reliable. From clinical experience of some Diabetes SIG members, it is very rare that someone has been diagnosed with diabetes and not know about it.
- [ARIC study has five diabetes definitions](#), and four of them use both medication and self-reported diagnoses information like the logic used in Definition 5 in HCHS/SOL.
- In HCHS/SOL, the Kappa agreement between self-reported medications and scanned medications is 0.96 (95% CI: 0.95-0.97).
- 

V1 Scanned medications	V1 Self-reported medications		Total
	No	Yes	
No	14,074	81	14,155
Yes	53	1,765	1,818
Total	14,127	1,846	15,973

**Concerns for including self-reported DIAGNOSIS:**

- Self-reported diabetes diagnosis might not be very reliable (i.e., some participants might mistakenly report being diabetic when they are not, and such misreporting could have significant impact on the answers to the research question). In HCHS/SOL at V2, there are a total of 644 participants who self-reported diagnosis and no medication use BUT with normal lab values (n=123) or impaired glucose tolerance lab values (n=521; see table below).



**Concerns for including self-reported MEDICATION USE:**

- Medication can alter lab measures, especially blood glucose. However, this concern may be mitigated somewhat because four weeks of treatment may not be sufficient to alter A1c as red blood cells have a lifespan of 90-120 days.
- Medications traditionally used for diabetes may be prescribed for an indication other than diabetes (e.g., metformin for PCOS or prediabetes, GLP-1 agonist for obesity, SGLT-2 inhibitor for cardiovascular disease and in near future probably CKD). However, this concern should be minimized given the wording of the question (MUE26c). Further, in HCHS/SOL at V2, there are only 41 participants who self-reported NO diabetes diagnosis BUT medication use with normal lab values (n=3) or impaired glucose tolerance lab values (n=38; see table below).

**Cross-classification of participants by lab values status, self-report medication use, and self-report diagnosis at Visit 2.**

LAB	Self-reported medication use	Self-reported diagnosis	Frequency
Normal	No	No	3084
	No	Yes	123
	Yes	No	3
	Yes	Yes	31
Impaired glucose tolerance	No	No	4481
	No	Yes	521
	Yes	No	38
	Yes	Yes	392
Diabetes	No	No	643
	No	Yes	278
	Yes	No	25
	Yes	Yes	1724
<b>TOTAL</b>			<b>11343</b>
<b>Highlighted DIABETES5 definition.</b>			

Highlighted in red for misclassification by DIABETES5\_V2 as a case by including self-reported medication use in the definition. Highlighted in blue for misclassification by DIABETES5\_V2 as a case by including self-reported diagnosis in the definition.