

## MINI-SENTINEL METHODS

### VALIDATING TYPE 1 AND TYPE 2 DIABETES MELLITUS IN THE MINI-SENTINEL DISTRIBUTED DATABASE USING THE SURVEILLANCE, PREVENTION, AND MANAGEMENT OF DIABETES MELLITUS (SUPREME-DM) DATALINK

**Prepared by:** Marsha A Raebel, PharmD,<sup>1,2</sup> Emily B Schroeder, MD, PhD,<sup>1,3</sup> Glenn Goodrich, MS,<sup>1</sup> Andrea R Paolino, MA,<sup>1</sup> William Troy Donahoo,<sup>3,4</sup> MD, Candace Fuller, PhD,<sup>5</sup> Carlos Bell, MPH,<sup>6</sup> Gregory A Nichols, PhD,<sup>7</sup> Julie Schmitt diel, PhD,<sup>8</sup> Katherine M. Newton, PhD,<sup>9</sup> Pamala Pawloski, PharmD,<sup>10</sup> Azadeh Shoaibi, PhD, MHS, MS<sup>6</sup>

**Author Affiliations:** 1. Kaiser Permanente Colorado, Institute for Health Research, Denver, CO, 2. Skaggs School of Pharmacy & Pharmaceutical Sciences, University of Colorado, Aurora, CO, 3. Department of Medicine, University of Colorado, Aurora, CO, 4. Colorado Permanente Medical Group, Denver, CO, 5. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim HealthCare Institute, Boston, MA, 6. Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, 7. Kaiser Permanente Northwest, Center for Health Research, Portland, OR, 8. Division of Research, Kaiser Permanente Northern California, Oakland, CA, 9. Group Health Research Institute, Seattle, WA, 10. HealthPartners Institute for Education and Research, Minneapolis, MN.

**April 8, 2016**

Mini-Sentinel is a pilot project sponsored by the [U.S. Food and Drug Administration \(FDA\)](#) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the [Sentinel Initiative](#), a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

## Acknowledgements

Additional Participating Members of the Validating Type 1 and Type 2 Diabetes Mellitus in the Mini-Sentinel Distributed Database using the SURveillance, PREvention, and ManagEMENT of Diabetes Mellitus (SUPREME-DM) DataLink Workgroup

Affiliation	Name
Institute for Health Research, Kaiser Permanente (KP) Colorado	John F Steiner
Institute for Health Research, KP Colorado	Nikki Carroll
Institute for Health Research, KP Colorado	Kristin Goddard
Center for Health Research, KP Northwest	Daniel S Sapp
Center for Health Research, KP Northwest	Xiuhai Yang
Center for Health Research, KP Northwest	Amanda Petrik
Division of Research, KP Northern California	Deanne Wiley
Division of Research, KP Northern California	Jack Hamilton
Group Health Research Institute	Monica Fujii
Group Health Research Institute	Roy Pardee
HealthPartners Institute for Education and Research	Jody Jackson
HealthPartners Institute for Education and Research	Terese DeFor
Center for Drug Evaluation and Research, United States Food and Drug Administration	Aaron Niman
Harvard Pilgrim Health Care Institute	Jennifer Popovic
Harvard Pilgrim Health Care Institute	Sandra Feibelmann
Harvard Pilgrim Health Care Institute	Pilar Williams
Harvard Pilgrim Health Care Institute	Bismark Baidoo
Harvard Pilgrim Health Care Institute	Casey Covarrubias
Harvard Pilgrim Health Care Institute	Ryan Saliga

## Mini-Sentinel Methods

### Validating Type 1 and Type 2 Diabetes Mellitus in the Mini-Sentinel Distributed Database Using the Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) Datalink

#### Table of Contents

<b>I. EXECUTIVE SUMMARY.....</b>	<b>- 1 -</b>
<b>II. INTRODUCTION AND SPECIFIC AIMS .....</b>	<b>- 3 -</b>
<b>III. THE SUPREME-DM DATALINK .....</b>	<b>- 5 -</b>
<b>IV. DATA PARTNERS AND DATA DEVELOPMENT .....</b>	<b>- 7 -</b>
A. DATA PARTNERS.....	- 7 -
B. DATA DEVELOPMENT .....	- 8 -
<b>V. SUPREME-DM DATALINK REFRESH.....</b>	<b>- 8 -</b>
<b>VI. SPECIFIC AIM 1: LINKING THE PATIENTS IN THE MSDD AND THE SUPREME-DM DATALINK.....</b>	<b>- 8 -</b>
A. DIRECT AND PROBABALISTIC MATCHING.....	- 8 -
B. FINAL PROJECT POPULATION SIZE.....	- 10 -
<b>VII. SPECIFIC AIM 2: DETERMINING THE SENSITIVITY (SE) AND POSITIVE PREDICTIVE VALUE (PPV) OF DIAGNOSIS CODES AND MEDICATION CLAIMS FOR DIABETES MELLITUS.....</b>	<b>- 10 -</b>
A. GENERAL APPROACH.....	- 10 -
B. SE AND PPV OF PUBLISHED ALGORITHMS FOR PATIENTS WITH ANY DM IN THE MSDD COMPARED TO THE SUPREME-DM DATALINK AS THE GOLD STANDARD .....	- 11 -
1. <i>Gold Standard for Any DM: SUPREME-DM DataLink Algorithm .....</i>	<i>- 13 -</i>
2. <i>Modified Gold Standard for Any DM: SUPREME-DM DataLink Algorithm without Laboratory Results Criteria .....</i>	<i>- 14 -</i>
3. <i>Impact on Case Ascertainment of Including Laboratory Test Results Criteria in the SUPREME-DM DataLink Gold Standard Algorithm .....</i>	<i>- 15 -</i>
4. <i>Solberg Primary Any DM Algorithm.....</i>	<i>- 16 -</i>
a. SE of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites .....	- 16 -
b. SE of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group .....	- 16 -

- c. SE of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM Algorithm DataLink Gold Standard by Gender ..... - 16 -
- d. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites ..... - 17 -
- e. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Age Group ..... - 17 -
- f. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Gender..... - 17 -
- 5. *Interpretation of the SE and PPV of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard* ..... - 17 -
- 6. *Solberg Secondary Any DM Algorithm (Solberg Modified to include Emergency Department (ED) Visits in Diagnosis Criteria)*..... - 17 -
  - a. SE of the Solberg Secondary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites ..... - 18 -
  - b. SE of the Solberg Secondary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Age Group ..... - 18 -
  - c. SE of the Solberg Secondary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Gender..... - 18 -
  - d. PPV of the Solberg Secondary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites ..... - 18 -
  - e. PPV of the Solberg Secondary *Any* DM Algorithm versus Gold Standard by Age Group.... - 19 -
  - f. PPV of the Solberg Secondary *Any* DM Algorithm versus Gold Standard by Gender..... - 19 -
- 7. *Interpretation of the SE and PPV of the Solberg Secondary Any DM Algorithm versus the Gold Standard*..... - 19 -
- 8. *Comparison of the Solberg Primary and Secondary Any DM Algorithms*..... - 19 -
- 9. *Zgibor Any DM Algorithm*..... - 20 -
  - a. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites..... - 20 -
  - b. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group ..... - 20 -
  - c. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Gender ..... - 21 -
  - d. PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites..... - 21 -
  - e. PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group..... - 21 -

f.	PPV of the Zgibor Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Gender .....	- 21 -
10.	<i>Interpretation of the SE and PPV of the Zgibor Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard</i> .....	- 21 -
11.	<i>Comparison of the Solberg Primary Any DM Algorithm, the Solberg Secondary Any DM Algorithm, and the Zgibor Any DM Algorithm versus the Gold Standard</i> .....	- 22 -
C.	COMPARISON OF PUBLISHED ALGORITHMS FOR PATIENTS WITH T1DM IN THE MSDD .....	- 22 -
1.	<i>Klompas Primary T1DM Algorithm</i> .....	- 24 -
a.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm across Sites Combined and by Individual Sites.....	- 24 -
b.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm by Age Group .....	- 24 -
c.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm by Gender .....	- 24 -
2.	<i>Klompas Optimized T1DM Algorithm</i> .....	- 25 -
a.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm across Sites Combined and by Individual Sites.....	- 25 -
b.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm by Age Group.....	- 25 -
c.	Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm by Gender .....	- 25 -
3.	<i>Comparison of Klompas Primary T1DM and Klompas Optimized T1DM Algorithm with Neither Considered a Gold Standard</i> .....	- 26 -
4.	<i>Interpretation of Case Identification using the Klompas Primary and Optimized T1DM Algorithms</i> .....	- 26 -
D.	CASE ASCERTAINMENT USING PUBLISHED ALGORITHMS FOR PATIENTS WITH T2DM IN THE MSDD .....	- 27 -
1.	<i>eMERGE Primary T2DM Algorithm</i> .....	- 29 -
a.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Primary Algorithm across Sites Combined and by Individual Site .....	- 29 -
b.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Primary Algorithm by Age Group .....	- 30 -
c.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by eMERGE Primary Algorithm by Gender .....	- 30 -
2.	<i>eMERGE Secondary T2DM Algorithm</i> .....	- 30 -
a.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm across Sites Combined and by Individual Site .....	- 30 -

b.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm by Age Group.....	31 -
c.	Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm by Gender .....	31 -
3.	<i>Comparison of the eMERGE Primary T2DM Algorithm and the eMERGE Secondary T2DM Algorithm with Neither Considered a Gold Standard</i> .....	31 -
4.	<i>Interpretation of Case Identification using the T2DM eMERGE Primary and Secondary Algorithms</i> .....	32 -
E.	PATIENTS WITH ANY DM WHO HAVE DM OF UNCERTAIN TYPE AFTER APPLYING T1DM AND T2DM ALGORITHMS-	32 -
1.	<i>Why Individuals Did Not Meet eMERGE T2DM Algorithm Criteria</i> .....	33 -
2.	<i>Why Individuals Did Not Meet the Klompas Optimized T1DM Algorithm Diagnosis Ratio Criteria</i> -	34 -
3.	<i>Exploration of Including C-Peptide and Diabetes Autoantibody Laboratory Test Results as Criteria for T1DM (Originally Pathways of Klompas Optimized Algorithm)</i> .....	35 -
<b>VIII.</b>	<b>RECOMMENDATIONS FOR USING THE SUPREME-DM DATALINK FOR ANY DM, T1DM, AND T2DM -</b>	<b>37 -</b>
<b>IX.</b>	<b>RECOMMENDED ALGORITHMS FOR IDENTIFYING ANY DM, T1DM AND T2DM IN THE MSDD POPULATION .....</b>	<b>38 -</b>
A.	IDENTIFYING ANY DM IN THE MSDD POPULATION .....	38 -
B.	CRITERIA/ALGORITHM(S) FOR IDENTIFYING T1DM IN THE MSDD POPULATION .....	39 -
C.	CRITERIA/ALGORITHM(S) FOR IDENTIFYING T2DM IN THE MSDD POPULATION .....	41 -
1.	<i>Recommended Option 1: Identifying T2DM in the MSDD Population</i> .....	42 -
2.	<i>Recommended Option 2: Identifying Non-T1DM in the MSDD Population</i> .....	43 -
D.	IMPLICATIONS FOR SENTINEL ROUTINE TOOLS .....	43 -
<b>X.</b>	<b>RECOMMENDATIONS FOR FUTURE WORK .....</b>	<b>44 -</b>
A.	FURTHER ENHANCEMENT OF THE SUPREME-DM DATALINK.....	44 -
1.	<i>Ongoing Refresh of the SUPREME-DM DataLink Database</i> .....	44 -
2.	<i>Test the Newly-developed T2DM Algorithm in the SUPREME-DM DataLink Any DM Population</i> -	44 -
B.	FURTHER IMPROVEMENTS TO THE SE AND PPV OF THE ALGORITHMS RECOMMENDED FOR USE IN THE MSDD POPULATION.....	44 -
C.	ADDING C-PEPTIDE AND DIABETES AUTOANTIBODIES LABORATORY TEST RESULTS TO THE MSDD TO IMPROVE CASE IDENTIFICATION OF ADULT-ONSET T1DM .....	47 -
D.	ALGORITHM DIAGNOSIS CODE CONVERSION TO ICD-10 .....	48 -

<b>XI. TABLES AND FIGURES .....</b>	<b>- 49 -</b>
A. TABLES .....	- 49 -
B. FIGURES.....	- 49 -
<b>XII. REFERENCES .....</b>	<b>- 50 -</b>
<b>XIII. APPENDIX A .....</b>	<b>- 55 -</b>
<b>XIV. APPENDIX B .....</b>	<b>- 65 -</b>

## I. EXECUTIVE SUMMARY

The FDA requested that we test an alternative strategy to validating diagnostic algorithms for diabetes mellitus (DM) that avoids medical record validation. The **SUR**veillance, **PRE**vention, and **ManagE**ment of **Diabetes Mellitus** (SUPREME-DM) DataLink (or registry) is a national distributed database of individuals with *Any DM* (i.e., Type 1 DM [T1DM], Type 2 DM [T2DM], and DM of rare or uncertain types). In this project, we employed data from five Mini-Sentinel Data Partners that are also SUPREME-DM DataLink sites.

We linked the Mini-Sentinel Distributed Database (MSDD) and the SUPREME-DM DataLink using variables found in both (probabilistic match) and using patient health record numbers maintained in DataLink and MSDD crosswalks at sites (direct match). With the direct approach, >99.9% of individuals in the SUPREME-DM DataLink were linked to individuals in the MSDD, while in the probabilistic approach, 98.8% were linked. **A total of 737,122 adults were linked** (01/01/2006 – 06/30/2014). **We conclude the SUPREME-DM DataLink is an excellent source registry for Any DM in adults.**

The SUPREME-DM DataLink *Any DM* Algorithm uses diagnoses, laboratory test results, and medication criteria. Considering the SUPREME-DM DataLink Algorithm the gold standard for *Any DM* in adults, we determined the sensitivity (SE) and positive predictive value (PPV) of other *Any DM* algorithms. To evaluate the impact of laboratory criteria on case ascertainment, we tested a modified SUPREME-DM DataLink Algorithm without laboratory criteria (i.e., diagnoses and medications only), identifying 5% fewer adults with *Any DM* than with the laboratory criteria. The workgroup tested three additional *Any DM* algorithms.<sup>1-3</sup> All had excellent SE ( $\geq 93.6\%$ ) and PPV ( $\geq 98.1\%$ ) relative to the SUPREME-DM DataLink Algorithm. **We conclude that all tested algorithms worked well at identifying Any DM.**

No gold standard algorithms for T1DM or T2DM are available for *adults*. Therefore, SE and PPV of existing T1DM or T2DM algorithms could not be determined. We did compare numbers and proportions of T1DM and T2DM cases identified in the MSDD using published algorithms. Two T1DM Algorithms<sup>4</sup> were tested. The T1DM algorithm that identified more cases across sites identified 3.6% (n = 26,418) of adults with *Any DM* as T1DM. This is below national estimates of T1DM in adults (4.6% - 4.8%).<sup>5</sup> While the workgroup considers the T1DM cases identified using the algorithm as true cases, there is clearly under-ascertainment with this algorithm. New-onset T1DM can be defined as the presence of diabetes antibodies or a low or negative C-peptide value. Because neither of these laboratory result types is in the MSDD, we had to modify the tested algorithm, omitting the laboratory criteria. To determine how case ascertainment changed when DM autoantibody and C-peptide results *were* included, we extracted these laboratory results from source data into the SUPREME-DM DataLink at one site. The proportion of adults identified with T1DM at that site increased 0.4% (from 4.8% to 5.2%), confirming the importance of T1DM laboratory results when identifying adults with T1DM. **We conclude case ascertainment of adult T1DM improves when the algorithm can include C-peptide and DM autoantibody criteria.**

Two T2DM Algorithms<sup>6</sup> were tested. The algorithm that identified more cases identified 70.7% of adults with *Any DM* as T2DM. *Any DM* cases identified as T2DM were even lower at sites without glucose-related laboratory results in the MSDD. **We conclude that the tested T2DM algorithms under-ascertain T2DM cases and cannot be recommended.**

Considering both the tested T1DM and T2DM algorithms, still > 25% of adults with *Any DM* were not identified as either T1DM or T2DM. We explored why individuals did not meet tested algorithm T1DM or T2DM criteria and incorporated those findings into recommendations.



### **Recommendations for using the SUPREME-DM DataLink**

- *Any DM*: Employ the SUPREME-DM DataLink as the gold standard registry for adults. With the linkage established between SUPREME-DM DataLink and MSDD, the SUPREME-DM DataLink can replace medical record review and be used as the alternative reference source for *Any DM*.
- *T1DM*: Consider the SUPREME-DM DataLink the gold standard registry for adults. We believe the tested *T1DM* algorithm identified a higher *number* of adults in the SUPREME-DM DataLink as *T1DM* than are available in any other *T1DM* registry. SUPREME-DM sites are adding DM autoantibody and C-peptide results to the DataLink; these laboratory results will further increase the number of true cases of adult *T1DM* available in the DataLink.
- *T2DM*:
  - Consider the SUPREME-DM DataLink the gold standard *non-T1DM* registry for adults, that is, for adults with *T2DM* and DM of uncertain and rare types. Nearly all adults that do not have *T1DM* have *T2DM*. For most Sentinel DM-related safety surveillance activities, the medical products of interest are used similarly in patients with *T2DM* and DM of uncertain or rare types. Differentiating between *T2DM* and DM of uncertain or rare types is usually unnecessary.
  - Prior to deeming the SUPREME-DM DataLink the gold standard *T2DM* registry for adults, consider testing a newly-developed *T2DM* Algorithm in the SUPREME-DM DataLink *Any DM* population. This algorithm retains most diagnosis and medication criteria from the tested *T2DM* algorithms and adds criteria to identify additional patients such as adults with *T2DM* not taking *T2DM* medication and who do not have laboratory results available. The workgroup believes this algorithm will identify 83% - 90% of adults in the SUPREME-DM DataLink as having *T2DM*.
- There are many advantages to employing the SUPREME-DM DataLink as the gold standard DM registry in medical product surveillance. Examples: 1) established linkage for > 737,000 adults; this number increases by > 50,000 adults yearly, 2) data in the SUPREME-DM DataLink are not consistently in the MSDD (e.g., additional laboratory result types, social behavioral data, race and ethnicity, cause of death), and 3) years of data are available for most adults.

### **Recommendations for using algorithms to identify adults with *Any DM*, *T1DM*, and *T2DM* in the MSDD**

- *Any DM*: Apply the SUPREME-DM DataLink Algorithm as the gold standard
- *T1DM*: First, individuals should meet *Any DM* SUPREME-DM DataLink algorithm criteria. Second, apply the Modified Klompas *T1DM* algorithm (without laboratory results). The limitation is under-ascertainment of cases because *T1DM* laboratory results are not available.
- *T2DM*: First, individuals should meet *Any DM* SUPREME-DM DataLink algorithm criteria. Second, apply Option 1 or 2.
  - Option 1 (*T2DM* algorithm): A newly-developed algorithm that retains most diagnosis and medication criteria from the tested *T2DM* algorithms and adds criteria to identify additional patients such as adults with *T2DM* not taking *T2DM* medication and who do not have laboratory results available. This algorithm minimizes case misclassification, but is somewhat complex to implement because of multiple criteria sets.
  - Option 2 (*non-T1DM* algorithm): Identify and exclude adults with *T1DM* using the Modified Klompas *T1DM* algorithm. Classify all remaining adults as *T2DM*. This Option is easy to implement, but misclassifies some cases because DM of uncertain and rare types are considered *T2DM*.

### **Implications for Sentinel Routine Query Tools**

- This workgroup focused on utilizing the SUPREME-DM DataLink (or registry) to explore algorithms to identify cohorts of individuals with ‘Any DM’, ‘T1DM’, and ‘T2DM’ within the Sentinel Distributed Database. Although workgroup aims were accomplished with *de novo* code, it would be possible to utilize Sentinel routine tools for algorithm implementation if recommended parameters are modified to include index dates. Sentinel routine tools were designed in the context of medical product safety surveillance, and require use of index dates to identify cohorts and health outcomes of interest. They utilize inclusion/exclusion criteria for cohort selection, which are assessed during a requester defined number of days before, on, or after the exposure episode index date. Similarly, Sentinel tools also require use of index dates to identify specific health outcomes of interest. Although the workgroup did not focus on identifying incident DM outcomes subsequent to a medical product exposure, it would be possible to modify recommended algorithm parameters (e.g., specify index dates) to capture Any DM and T2DM outcomes.
- Algorithms for Any DM and T2DM may be able to be implemented in the current Sentinel tools if all parameters and temporal relationships are defined in a tool-specific manner. For example, as algorithms for Any DM exclude pregnancy, an already developed algorithm for pregnancy would need to be adapted to the Sentinel tool framework. NDC lists for relevant antidiabetic agents have already been developed by the Workgroup and, similar to use of other NDC lists, would need periodic updates. All examined DM algorithms include several criteria (or criteria sets), and temporal relationships are at times, integral. Thus, timeframes included within each algorithm would need to be reinterpreted in terms of index dates. Current Sentinel tools are able to utilize the glucose and glycosylated hemoglobin (HbA1c) laboratory result values included in the Sentinel Common Data Model; it is possible to distinguish elevated versus non-elevated random and fasting glucose values and HbA1c at Data Partners that contribute laboratory result values to the laboratory results table. Thus, the individual criteria for Any DM and T2DM (if reinterpreted) appear to be compatible with Sentinel tools.
- To implement algorithms for T1DM examined in this report, Sentinel tools would need to be modified to accommodate ratios of T1DM to T2DM codes. However, should accommodation of these T1DM algorithms be deemed a priority, it would be possible to determine if the tools can be updated to accommodate these algorithms.
- In summary, algorithms for Any DM and T2DM described in this report appear to be compatible with current Sentinel tools, but algorithms for T1DM currently cannot be implemented using existing Sentinel tools. The Sentinel Operations Center recommends piloting algorithm implementation to confirm the Any DM and T2DM algorithm parameters can be implemented as intended using the existing query tools. This will assist with identifying any gaps between current and needed tool functionality. Careful consideration of potential medical product safety question (s) of interest during this pilot will also help to ensure algorithms are implemented appropriately.

## **II. INTRODUCTION AND SPECIFIC AIMS**

Diabetes mellitus (DM) is a high priority health outcome of interest (HOI) to the FDA. However, most studies that have validated DM algorithms did not distinguish between T1DM and T2DM in adults. Further, accurate identification of Any DM can be problematic in the Mini-Sentinel Distributed Database

(MSDD) because electronic databases rely on ICD codes alone or in combination with data algorithms to identify outcomes. Misclassification is always a concern with electronic observational databases, and the “gold standard” approach is to employ algorithms validated using full-text medical record review. Whether or not existing DM algorithms used data or populations similar to those in the MSDD is not known. Additionally, resources required for extensive medical record validation can be substantial.

The Mini-Sentinel Alternative Methods for HOI Validation workgroup summarized alternative methods for validating selected HOIs, including linking to registries and external data sources, as compared to chart validation ([http://mini-sentinel.org/methods/outcome\\_validation/details.aspx?ID=105](http://mini-sentinel.org/methods/outcome_validation/details.aspx?ID=105)).<sup>7</sup> In that project, methods that are less costly and time-consuming than medical record review were explored. The HOIs selected in that project were based on a dearth of validated algorithms with high performance characteristics in the existing literature in addition to FDA priority. DM was not only rated as a high priority HOI, but was also considered feasible for linkage to a registry based on the availability and accessibility of alternate data sources, cost of linking alternative data sources to the MSDD, and the extent of overlap between available alternative databases and the MSDD.

The **SUR**veillance, **PRE**vention, and **ManagE**ment of **Diabetes Mellitus** (SUPREME-DM) DataLink is a distributed database (or registry) developed for studying *Any* DM in usual care environments in the United States. It brings together extensive data from electronic health records (EHR) and other clinical and administrative databases from 11 representative healthcare delivery and insurance systems, nine of which also participate in Mini-Sentinel. Detailed information about development and implementation of the SUPREME-DM DataLink is provided in Section III. In brief, patients in the SUPREME-DM DataLink were identified from combinations of diagnoses, laboratory test results, and medication use found in electronic health record (EHR) and other clinical and administrative electronic data. At several sites, the DM cases in the SUPREME-DM DataLink are patients in site-specific DM registries. Through 2012, the SUPREME-DM DataLink represented a defined population of over 1.1 million adult patients across the United States with *Any* DM, including over 250,000 with incident DM and an estimated 40,000 to 50,000 patients with T1DM. The proportions of adult patients in the SUPREME-DM DataLink with T1DM and with T2DM approximate the proportions in the United States adult population.

This current Mini-Sentinel workgroup was charged with employing an alternative strategy to validating a diagnostic algorithm that avoids extensive medical record validation. Specifically, the FDA asked this workgroup to test this alternative strategy using the SUPREME-DM DataLink electronic database registry that contains true cases of *Any* DM, and to link it to the MSDD. In this project, we investigated the feasibility of this alternative validation method using the SUPREME-DM DataLink data from five Mini-Sentinel Data Partner sites that are also SUPREME-DM sites to undertake the following:

- Specific Aim 1: Demonstrate the feasibility of linking the MSDD and the SUPREME-DM DataLink. Because participating SUPREME-DM DataLink sites are also MSDD Data Partner sites, we anticipated linkage would approach 100%. Therefore, the main activity within this Aim was to develop and implement the data table cross-walk that identified and linked patients in the MSDD and the SUPREME-DM DataLink in the distributed data environment.
- Specific Aim 2: Determine the SE and PPV of diagnosis codes and medication claims for patients with a) *Any* DM (e.g., inclusive of T1DM, T1DM, and DM of uncertain of rare types), b) T1DM, and c) T2DM present in the MSDD, compared to the SUPREME-DM DataLink algorithm as the gold standard algorithm for *Any* DM.

The workgroup suggested conducting a Specific Aim 3. The objective of Aim 3 would have included reviewing a sample of medical records from patients identified in Aim 2 as, for example, having DM of

uncertain type (e.g., patients with mixed T1DM and T2DM coded diagnoses). The workgroup also anticipated that the criteria sets applied in Aim 2 likely would not perform in the MSDD population as they did in the populations where they had been originally studied (e.g., because they were developed in a much smaller population than that in the MSDD, had not been previously externally validated, or required data elements not available in the MSDD). Reviewing medical records as part of this project would assist in estimating validity of tested criteria sets and allow the workgroup to determine characteristics associated with discrepancy, ultimately informing an approach to reduce uncertainty (e.g., algorithm refinement). Reviewing a sample of medical records within the context of the current project could also help avoid the need to justify foregoing medical record review in the future. However, Aim 3 was deemed outside the scope of the current project because this project was intended to be incremental towards the larger objective of alternative HOI validation.

### III. THE SUPREME-DM DATALINK

Registries of patients with DM have been available for decades. Older DM registries were single-site registries comprised of coded diagnoses and medication prescription data from administrative claims. Recently, the availability of detailed clinical data from sources such as EHR and laboratory test results databases enabled building DM registries that included more complete DM cohorts and detailed patient information. Newer registries are still usually single-site registries. While several single-site studies have validated administrative definitions of DM,<sup>1,8-10</sup> criteria for inclusion in single-site DM registries lack uniformity, rendering across-registry comparisons of limited usefulness.<sup>11</sup> Further, single-site registries typically include fewer than 100,000 individuals with DM and often contain just a few thousand patients. Single-site registries are of insufficient size to be useful for many comparative effectiveness and safety activities.

The lack of uniform criteria and data elements across existing single-site DM registries and the relatively small numbers of patients with DM in single-site registries compelled the development of a robust multi-site *Any* DM registry. In 2010, funding from the Agency for Healthcare Research and Quality (AHRQ; grant number R01HS019859) under the PROSPECT (Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests) initiative established the SUPREME-DM Network, which was comprised of 11 integrated healthcare delivery systems (<http://www.supreme-dm.org/>). The SUPREME-DM DataLink employed a standardized methodology for identifying people with DM using the detailed clinical information available in EHRs and applied that methodology across the multiple health systems. Many sites and researchers that participated in developing the SUPREME-DM DataLink had previous experience developing and using single-site DM registries.<sup>8,12-15</sup> The success of the SUPREME-DM DataLink is evident: since 2010, a series of papers have been published that emphasize DM surveillance, disparities, complications, and pharmacotherapy.<sup>16-27</sup>

The SUPREME-DM consortium brought together nearly three dozen DM researchers from 11 organizations that are members of the Health Care Systems Research Network (HCSRN, formerly the HMO Research Network). Organizations that participate in SUPREME-DM include six Kaiser Permanente regions (Northern California, Southern California, Northwest [Oregon/Washington], Hawaii, Colorado, and Georgia), HealthPartners (Minnesota), Marshfield Clinic (Wisconsin), Geisinger Health System (Pennsylvania), Group Health Cooperative (Washington), and Henry Ford Health System (Michigan). Five of these SUPREME-DM sites participated in this current Mini-Sentinel project (Section IV).

Developing the SUPREME-DM DataLink multi-site registry required a different approach than a single-site registry. As recommended by Richesson,<sup>28</sup> a standardized methodology was employed across all sites to build and maintain the SUPREME-DM DataLink in a manner that ensured data could be appropriately aggregated across sites and organizations. For example, identical data variable definitions were employed, data were extracted into similarly formatted data tables at each site, and routine data quality checks were conducted on the diverse population data from the multiple sites and healthcare delivery systems represented in SUPREME-DM.

Building the SUPREME-DM DataLink benefitted from using the existing HCSR Virtual Data Warehouse (VDW), a data resource supported by member organizations and network consortia.<sup>29</sup> Within each participating site, when building the VDW, data were extracted from health plan databases and configured into identically-formatted VDW tables using standard variable names and values. The VDW data tables used for SUPREME-DM are far more comprehensive than the data available to commercial insurers, in Medicare or Medicaid databases, or in the MSDD because VDW tables include extensive EHR data as well as administrative claims data. The data can be linked through a common unique patient identifier. This approach enables use of standardized data extraction programs distributed to all sites. Each site constructs individual-level data sets for analysis and either a) sends limited or de-identified data sets to the lead site where they are combined for analysis, or b) retains the individual-level data sets at individual sites and run distributed analytic programs on their site's project-specific data set. Nine SUPREME-DM sites use an EPIC-based EHR (EPIC, Verona, Wisconsin). Examples of VDW data that are also available in the SUPREME-DM DataLink include race and ethnicity, comprehensive laboratory result values (e.g., chemistry laboratory test results [glycosylated hemoglobin, fasting and random plasma glucose, serum lipids, electrolytes, thyroid function, liver enzymes] and hematology and coagulation laboratory test results, vital signs (blood pressure, height, weight), and social history (tobacco use, substance use). All data can be used to assess the presence and intensity of conditions such as hyperglycemia, dyslipidemia, hypertension, and obesity. At all participating sites, at least 90% of members have a pharmacy benefit that helps ensure near complete identification of medication dispensings.<sup>16</sup>

DM indicator variables used to develop the initial SUPREME-DM DataLink included inpatient or outpatient DM diagnoses, glucose-related laboratory test results, and medication dispensings available from the EHR, clinical, and administrative databases of the health systems. These indicator variables were applied in an algorithm to identify individuals with *Any* DM (detailed information about the SUPREME-DM DataLink Algorithm is in Section VII.B).<sup>16</sup> Patients with DM are recognized in a reasonably short period of time using the SUPREME-DM DataLink algorithm, in part because multiple data sources and variable types are employed. When the SUPREME-DM DataLink was developed, about 85% of DM cases identified at all sites had multiple DM indicators.

The initial iteration of the SUPREME-DM DataLink included nearly 1.1 million members of all ages with DM from these 11 health systems between 2005 and 2009. On average, these individuals had 5 years of health system membership after DM identification. Mean age at DM identification (55.7 years) and the proportion of women (48.1%) were consistent across systems.<sup>16</sup> Data from 2010 through 2012 were later added to the SUPREME-DM DataLink, bringing the number of individuals in the SUPREME-DM DataLink registry to over 1.3 million.

The SUPREME-DM DataLink capitalized on strengths of the participating organizations, such as defined populations and rich longitudinal data. As Nichols et al pointed out in 2012, the SUPREME-DM DataLink is unique in size, comprehensiveness, and geographic coverage.<sup>16</sup> Other previous or existing electronic DM registries cannot provide data for analysis equivalent to the SUPREME-DM DataLink because they



have not been maintained,<sup>30</sup> are comprised of non-representative populations,<sup>31</sup> or only include patients from a single geographic region.<sup>3</sup>

Use for safety surveillance is a goal of SUPREME-DM. The efficiency of the DataLink and SUPREME-DM network was initially tested in 2012. Judith Fradkin, MD, Director, Division of Diabetes, Endocrinology, & Metabolic Diseases at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) inquired about SUPREME-DM's capability to answer questions about the risks of pancreatic cancer among patients with DM treated with dipeptidyl peptidase IV (DPP-IV) Inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists. The program to extract information about the number of individuals exposed to these therapies was distributed and data from 10 sites were received, tabulated, and provided back to NIDDK < 3 working days later. Although the number of individuals exposed to these drugs at that time was small, the capability to efficiently identify medication exposures and to aggregate and report multi-site data using the SUPREME-DM DataLink registry was demonstrated.

Although the SUPREME-DM DataLink sites built and maintained the SUPREME-DM DataLink using identical methods and variables, site-specific variations can exist. Differences in how providers' code DM diagnoses or incomplete data capture could introduce site-specific variation in DM identification. Although a common case identification algorithm was used to identify members with DM across all SUPREME-DM sites, resources were not available to validate each of the 1.1 million cases using medical record review. Finally, the SUPREME-DM DataLink was developed as an *Any DM* registry. Work with the SUPREME-DM DataLink to date has not focused on distinguishing adults with T1DM versus T2DM.

Overall, the SUPREME-DM DataLink is the most robust and most representative EHR-based multi-site DM registry in the United States. It is a large and unique registry that provides opportunities to conduct multiple types of data for DM research and surveillance. Pertinent to the current Mini-Sentinel linkage and alternative validation project, the SUPREME-DM DataLink was chosen as the gold standard *Any DM* registry because it was a valuable, emerging resource for safety and effectiveness studies and related epidemiologic surveillance and research. If the T1DM and T2DM algorithms tested in this project identified substantial numbers of patients with T1DM or T2DM in the SUPREME-DM DataLink registry, it could also be explored as the potential gold standard for T1DM and T2DM registries.

## IV. DATA PARTNERS AND DATA DEVELOPMENT

### A. DATA PARTNERS

Five Mini-Sentinel Data Partners that are also SUPREME-DM DataLink sites participated in this project:

1. Institute for Health Research, Kaiser Permanente (KP) Colorado (KPCO), Denver, Colorado (lead site)
2. HealthPartners Institute for Education and Research, Minneapolis, Minnesota
3. Division of Research, KP Northern California, Oakland, California
4. Center for Health Research, KP Northwest, Portland, Oregon
5. Group Health Research Institute, Seattle, Washington

All five sites participated in Aims 1 and 2.

## B. DATA DEVELOPMENT

KPCO wrote distributed programming code to update the SUPREME-DM DataLink at the participating sites, to link the patients in the SUPREME-DM DataLink with the patients in the MSDD, and to execute the DM algorithms. The code was tested and quality-checked in accordance with Mini-Sentinel Principles and Policies.<sup>32</sup> All distributed code was sent to the participating Data Partners according to Mini-Sentinel Principles and Policies.<sup>32</sup> Data partner sites executed the work plans against their SUPREME-DM DataLink (data refresh, linking programs and DM algorithms) and MSDD (linking programs and DM algorithms) and retained their site-specific linked datasets. Results describing site-specific linkages were returned to the Mini-Sentinel Operations Center and KPCO. Site-specific DM algorithm datasets were returned to the Mini-Sentinel Operations Center and KPCO where the datasets from the five sites were combined to yield one analytic dataset for DM algorithm testing.

## V. SUPREME-DM DATALINK REFRESH

Through 2012, the five participating sites had an enrolled population of 10,900,772 individuals, including 722,069 patients with *Any* DM in the SUPREME-DM DataLink. Refreshing the SUPREME-DM DataLink with data from January 1, 2013 through June 30, 2014 added 1,095,612 patients (total n = 11,996,384) and an additional 83,640 patients with *Any* DM. Thus, the total number of patients with *Any* DM in the SUPREME-DM DataLink at these five sites was 805,709.

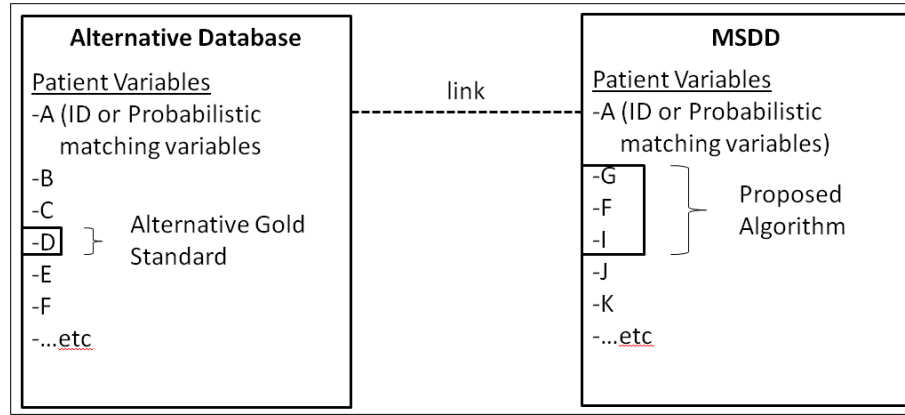
The date range of this project included January 1, 2006 through June 30, 2014. However, the date range of patients in the SUPREME-DM DataLink was broader (January 1, 2005 through June 30, 2014). After restricting to patients who entered the DataLink from January 1, 2006 through June 30, 2014, the total number of patients of any age with *Any* DM in the SUPREME-DM DataLink eligible for linking with the MSDD was 776,125.

## VI. SPECIFIC AIM 1: LINKING THE PATIENTS IN THE MSDD AND THE SUPREME-DM DATALINK

### A. DIRECT AND PROBABALISTIC MATCHING

The Alternative Methods for HOI Validation workgroup recommended validation processes of linking by patient identifier (direct matching) and by probabilistic matching (Figure 1).<sup>7</sup> In this project we demonstrated the feasibility of linking the MSDD and the SUPREME-DM DataLink both by patient identifier and by probabilistic matching by developing and implementing the necessary data table crosswalks.

**Figure 1. Linked HOI alternative validation<sup>a</sup>**



The Mini-Sentinel Data Partners participating in this project maintain patient identifier crosswalks at their local sites that enabled linking the unique patient identifier in the MSDD to the patient’s actual health record number and the unique patient identifier in the SUPREME-DM DataLink to the patient’s actual health record number.<sup>b</sup> Thus, patients in the SUPREME-DM DataLink could be linked to patients in the MSDD using variables found in both the SUPREME-DM DataLink and MSDD (probabilistic match) or using actual patient health record numbers maintained in SUPREME-DM DataLink and MSDD crosswalks at local sites (direct match).

Employing the probabilistic match potentially offers a durable method should future linkage be desirable to an additional patient registry (e.g., a registry where a direct match is not possible). Employing the direct match method enabled determining the accuracy of the probabilistic match and ensured near-complete linkage of the individuals in the SUPREME-DM DataLink to the individuals in the MSDD. Individuals matched in the direct match were considered the project population.

Both probabilistic and direct matching were conducted at the local sites and the resulting matched datasets were retained at each site. Only the denominator (total patients in that site’s SUPREME-DM DataLink) and the numerator (total patients matched at the site to the MSDD) were returned to the lead site.

In the direct match approach, > 99.9% of individuals were matched (776,037 of 776,125).

To be considered a match in the probabilistic approach, all of the following criteria must have matched:

- Total number of days enrolled
- Gender
- Date of birth
- The first two pharmacy dispensing dates and NDC's associated with them

<sup>a</sup> Reproduced from the final report of the Alternative Methods for Health Outcomes of Interest Validation Workgroup ([http://mini-sentinel.org/methods/outcome\\_validation/details.aspx?ID=105](http://mini-sentinel.org/methods/outcome_validation/details.aspx?ID=105))

<sup>b</sup> The actual health record number of each patient is not contained within either the MSDD or the SUPREME-DM DataLink



- First outpatient visit
- First inpatient visit

In the probabilistic approach, 98.8% of individuals were matched (766,999 of 776,125).

As shown in Table 1, the SE of the probabilistic match could be determined. The SE measures how well the probabilistic approach correctly linked individuals (calculated by dividing the number of correctly linked individuals by the total of correctly linked individuals plus “incorrectly not linked” individuals [additional individuals linked only with direct match]). The SE of the probabilistic match was 98.8% (766,998/776,037).

**Table 1. Probabilistic Match and Direct Match Gold Standard in Mini-Sentinel Distributed Database and SUPREME-DM DataLink Linkage Project**

Probabilistic Match	Gold Standard: Direct Match		Total
	Linked (Yes)	Not Linked (No)	
Linked (Yes)	766,998 Correctly Linked (True Positive; TP)	1 Incorrectly Linked (False Positive; FP)	766,999
Not Linked (No)	9,039 Incorrectly Not Linked (False Negative; FN)		
Total	776,037		

## B. FINAL PROJECT POPULATION SIZE

This project focused on adults. There were 38,915 linked individuals who were < 20 years of age. Therefore, limiting the individuals linked in the direct match to individuals aged 20 and older at the time of cohort entry yielded the final project population size of 737,122. These 737,122 adults aged 20 and older were included in validating the DM algorithms.

## VII. SPECIFIC AIM 2: DETERMINING THE SENSITIVITY (SE) AND POSITIVE PREDICTIVE VALUE (PPV) OF DIAGNOSIS CODES AND MEDICATION CLAIMS FOR DIABETES MELLITUS

### A. GENERAL APPROACH

The SUPREME-DM DataLink is the gold standard for identifying patients with *Any DM*.<sup>2,16,18,20,22,33,34</sup> SUPREME-DM DataLink patients were identified from combinations of diagnoses, laboratory test results, and in some situations, medication use criteria. The initial plan for this project was to use several criteria sets, including two published algorithms,<sup>4,35</sup> to identify patients with *Any DM* as well as to identify patients with T1DM, T2DM, and DM of uncertain type (e.g., mixed T1DM and T2DM codes) in the MSDD. We intended to determine the SE and PPV of sets of criteria (e.g., diagnosis codes only, combinations of coded diagnoses and oral or non-insulin injectable antidiabetic medication dispensings, combinations of coded diagnoses and only insulin dispensings) and the published algorithms and compare the case definitions. For each tested criteria set/algorithm, the patients identified with DM in the MSDD were to

be compared to the SUPREME-DM DataLink as the gold standard, particularly the new standard for validated T1DM case definition.

While the planned approach was in general maintained, the workgroup (including FDA investigators) identified several published DM algorithms<sup>1,3,4,6,8,35-40</sup> for *Any* DM and for T1DM and T2DM and determined that it would be more efficient and informative to test more than two published algorithms rather than develop new criteria sets. Details about each of the published algorithms the workgroup considered are in Appendix A. From these published algorithms, we selected algorithms that were most appropriate for testing in the MSDD population. For example, we required that the algorithm had been applied to adults. We also considered recommendations from the Mini-Sentinel Methods: 15 Cohorts of Interest for Surveillance Preparedness<sup>2</sup> workgroup, and gave preference to algorithms that had been externally validated. This process yielded nine algorithms for testing. These included:

- *Any* DM Algorithms
  - SUPREME-DM DataLink as the gold standard Algorithm<sup>2,16,18,20,22,33,34</sup>
  - SUPREME-DM DataLink Algorithm without laboratory results criteria. The Gold Standard Algorithm was revised to exclude laboratory results criteria to evaluate the impact of laboratory criteria on case ascertainment
  - Solberg Primary Algorithm<sup>1</sup>
  - Solberg Secondary Algorithm (as modified by the Mini-Sentinel Methods: 15 Cohorts of Interest for Surveillance Preparedness workgroup)<sup>1,2</sup>
  - Zgibor Algorithm<sup>3</sup>
- T1DM Algorithms
  - Klompas Primary Algorithm<sup>4</sup>
  - Klompas Optimized Algorithm (as modified by the Workgroup to align with the data elements available in the MSDD)<sup>4</sup>
- T2DM Algorithms
  - electronic MEdical Records and GENomics (eMERGE)<sup>6</sup> Primary Algorithm (without laboratory test results)
  - eMERGE Secondary Algorithm<sup>6</sup> (with laboratory test results)

As discussed later in this report, neither of the T1DM or T2DM Algorithms we tested is a gold standard. As a result, SE and PPV are not estimated for the T1DM and T2DM Algorithms we tested.

## **B. SE AND PPV OF PUBLISHED ALGORITHMS FOR PATIENTS WITH ANY DM IN THE MSDD COMPARED TO THE SUPREME-DM DATALINK AS THE GOLD STANDARD**

All of the *Any* DM algorithms include diagnosis and medication criteria (Table 2 and Appendix B). The SUPREME-DM DataLink algorithm (the gold standard) and the Zgibor et al *Any* DM algorithm also include laboratory test results. The modified SUPREME-DM DataLink *Any* DM algorithm does not include laboratory test results, nor does the primary Solberg *Any* DM algorithm or the modified Solberg secondary *Any* DM algorithm. Comparing the gold standard SUPREME-DM DataLink Algorithm with the modified SUPREME-DM DataLink Algorithm without laboratory test results at the participating sites with laboratory results in the MSDD enabled us to estimate the impact on these algorithms of not having laboratory test results available. As indicated in Table 2, prior to testing, some algorithms were updated to include newer diagnosis codes (i.e., available only after the date the algorithm was published) and

other minor modifications (e.g., to align with current rather than previous published DM guidelines) that did not alter the intent of the approach used in the original algorithm.

**Table 2. Any Diabetes Algorithms**

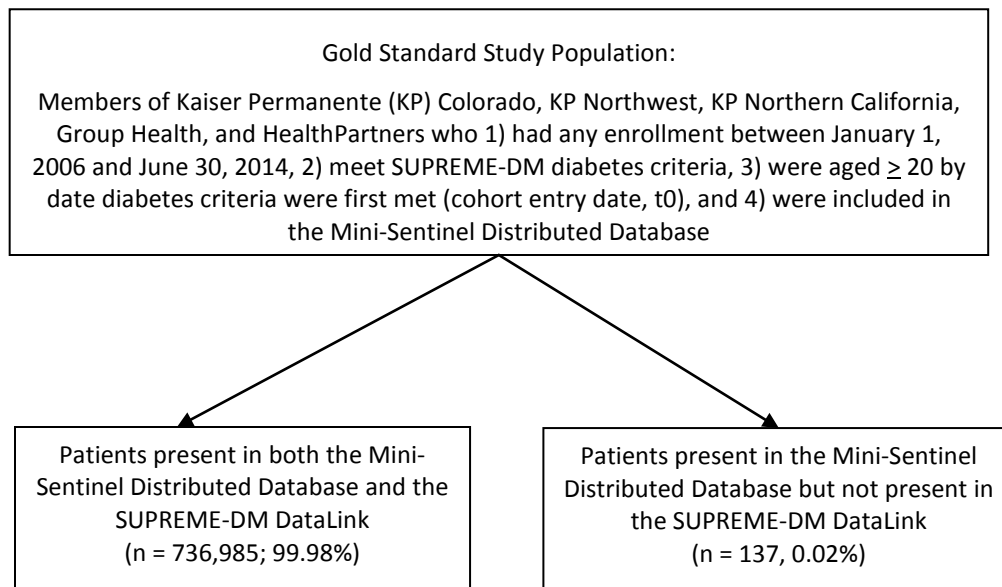
Algorithm Identifier	Original Algorithm	Tested/Modified Algorithm	Notes from 15 Cohorts WG <sup>2</sup>
<p><u>1. Gold Standard: "Any Diabetes" Algorithm: SUPREME-DM DataLink</u></p> <p>SUPREME-DM<sup>16-24,33,34</sup></p> <p>Diagnosis codes, medications, and labs</p>	<p>≥ 1 inpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41, OR</p> <p>Two of the following (when the two events were from the same source [e.g. two outpatient diagnoses or two elevated laboratory values], they must occur on separate days no more than 730 days apart):</p> <ol style="list-style-type: none"> <li>1) Outpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41</li> <li>2) Prescription for an antidiabetic medication (two dispensing of metformin or two dispensings of thiazolidinediones with no other indication of diabetes were not included)</li> <li>3) A1c ≥ 6.5%</li> <li>4) Fasting plasma glucose ≥ 126 mg/dl</li> <li>5) Random plasma glucose ≥ 200mg/dl</li> </ol> <p>Criteria ascertained during periods of pregnancy were excluded to ensure gestational diabetes was not inadvertently captured. The approach for all of the DM algorithms for periods of pregnancy will mirror that used for SUPREME-DM</p>	Original algorithm	N/A
<p><u>2. SUPREME-DM "Any Diabetes" Algorithm Without Labs</u></p> <p>Diagnosis codes and medications</p> <p>Nichols et al<sup>16</sup></p>	<p>≥ 1 inpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41, OR</p> <p>Two of the following:</p> <ol style="list-style-type: none"> <li>1) Outpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41</li> <li>2) Prescription for an antidiabetic medication (two dispensing of metformin or two dispensings of thiazolidinediones with no other indication of diabetes were not included)</li> </ol> <p>Criteria ascertained during periods of pregnancy were excluded to ensure gestational diabetes was not inadvertently captured. The approach for all of the DM algorithms will mirror that used for SUPREME-DM</p>	SUPREME-DM algorithm modified to remove the laboratory criteria (for the purposes of evaluating the impact of laboratory criteria on case ascertainment)	N/A
<p><u>3. Primary "Any Diabetes" Algorithm</u></p> <p>Diagnosis codes and</p>	<ol style="list-style-type: none"> <li>1) ≥ 2 outpatient ICD-9 codes from among the following, in a given calendar year: 250.XX, 357.2, 362.01, 362.02, 366.41, OR</li> <li>2) ≥ 1 inpatient ICD-9 codes from among the</li> </ol>	"In a given calendar year" was modified to "no more than 730 days apart"	Workgroup recommended for the primary "any diabetes"

Algorithm Identifier	Original Algorithm	Tested/Modified Algorithm	Notes from 15 Cohorts WG <sup>2</sup>
<p>medications</p> <p>Solberg<sup>1</sup></p>	<p>following, in a given calendar year: 250.XX, 357.2, 362.01, 362.02, 366.41, OR</p> <p>3) <math>\geq 1</math> prescription for an antidiabetic medication (excluding single-agent metformin) in a given calendar year</p>	<p>The list of retinopathy codes (362.01 and 362.02) in Solberg will be expanded to include codes 362.01-362.07. We believe that these additional retinopathy codes were added to the ICD-9-CM since the Solberg algorithm was developed.</p>	<p>algorithm</p>
<p><u>4. Secondary “Any Diabetes” Algorithm</u></p> <p>Diagnosis codes and medications</p> <p>Solberg<sup>1</sup></p>	<p>1) <math>\geq 2</math> outpatient ICD-9 codes from among the following, in a given calendar year: 250.XX, 357.2, 362.01, 362.02, 366.41, OR</p> <p>2) <math>\geq 1</math> emergency department/inpatient ICD-9 codes from among the following, in a given calendar year: 250.XX, 357.2, 362.01, 362.02, 366.41, OR</p> <p>3) <math>\geq 1</math> prescription for an antidiabetic medication (excluding single-agent metformin) within <math>\pm 365</math> days only if no ICD-9 diagnosis of 251.8, 256.4, or 962.0 occurs in the same calendar year or year prior</p>	<p>“In a given calendar year” was modified to “no more than 730 days apart”</p> <p>The list of retinopathy codes (362.01 and 362.02) in Solberg was expanded to include codes 362.01-362.07. We believe that these additional retinopathy codes were added to the ICD-9-CM after the Solberg algorithm was developed</p>	<p>Workgroup modification to Solberg et al; use in parallel with the primary recommendation to examine alternate definition if there is concern that the primary might miss persons diagnosed with diabetes in the emergency department</p>
<p><u>5. “Any Diabetes” Algorithm</u></p> <p>Diagnosis codes, medications, and labs</p> <p>Zgibor<sup>3</sup></p>	<p>Two or more of the following:</p> <ol style="list-style-type: none"> <li>1) ICD-9 code for 250.XX on an inpatient claim</li> <li>2) ICD-9 code for 250.XX on an outpatient claim</li> <li>3) ICD-9 code for 250.xx on an emergency department claim</li> <li>4) Prescription for an antidiabetic medication</li> <li>5) Any A1c measurement, regardless of value</li> <li>6) Blood glucose &gt; 200mg/dl</li> </ol> <p><b>OR</b> any single ICD-9 code for 250.XX on an outpatient claim</p>	<p>The SUPREME-DM algorithm applies random glucose <math>\geq 200</math> mg/dl (as well as fasting glucose <math>\geq 126</math> mg/dl and A1c <math>\geq 6.5\%</math>) in keeping with the 2011 and 2015 ADA DM diagnosis criteria.<sup>41,42</sup> The Zgibor criteria apply random glucose &gt; 200 mg/dl. We modified the Zgibor algorithm to apply random glucose <math>\geq 200</math> mg/dl. Because A1c is now used for diagnosis (which was not the case when the Zgibor algorithm was developed), changed the “any A1c measurement, regardless of value” criterion to an A1c <math>\geq 6.5\%</math>.</p>	<p>Zgibor et al; use if requisite laboratory data become widely available within the distributed database</p>

## 1. Gold Standard for Any DM: SUPREME-DM DataLink Algorithm

As shown in Figure 2, essentially 100% (99.98%; 736,985 of 737,122) of patients in the population in the MSDD were linked to patients in the SUPREME-DM DataLink using the gold standard SUPREME-DM DataLink diagnosis codes, medications, and laboratory test result algorithm criteria.

**Figure 2. SUPREME-DM DataLink Linkage to Mini-Sentinel Distributed Database Project Cohort for SUPREME-DM DataLink Algorithm Gold Standard for Any Diabetes**

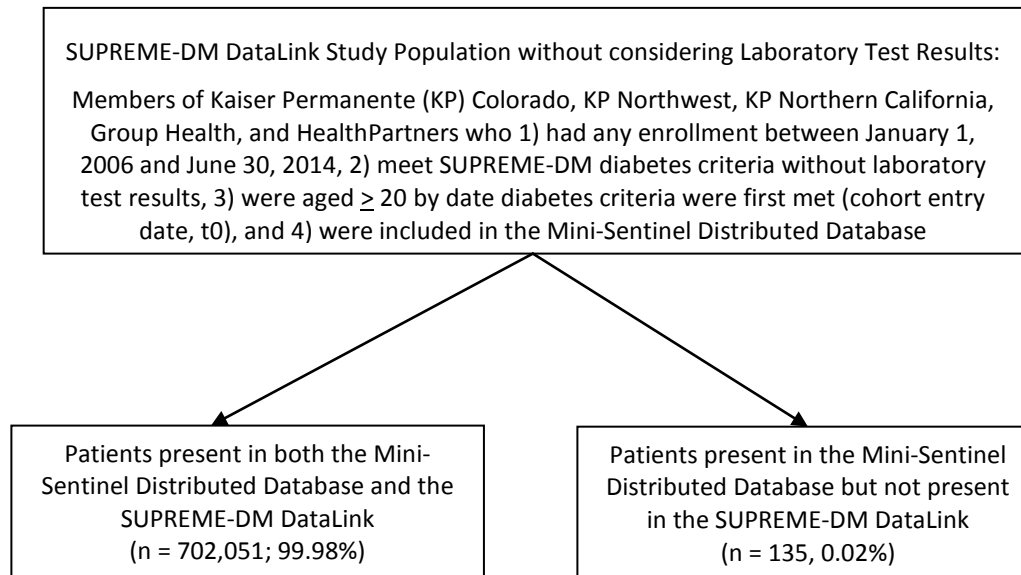


Only 137 (0.02%) patients were present in the MSDD but not in the SUPREME-DM DataLink. It is likely these represent data entry errors in the source databases (e.g., inconsistent health record numbers).

## 2. Modified Gold Standard for Any DM: SUPREME-DM DataLink Algorithm without Laboratory Results Criteria

As shown in Figure 3, essentially 100% (99.98%; 702,051/702,186) of patients in the gold standard population (in the MSDD) were linked to patients in the SUPREME-DM DataLink using the SUPREME-DM DataLink Algorithm without laboratory test results (medications and diagnoses only).

**Figure 3. Linkage to Mini-Sentinel Distributed Database Project Cohort for SUPREME-DM DataLink Any Diabetes Algorithm without Laboratory Results**



Only 135 (0.02%) patients were present in the MSDD but not in the SUPREME-DM DataLink. As previously noted, it is likely these 135 patients represent data entry errors in the source databases (e.g., inconsistent health record numbers).

### 3. Impact on Case Ascertainment of Including Laboratory Test Results Criteria in the SUPREME-DM DataLink Gold Standard Algorithm

At the time this work was conducted, four of the five Mini-Sentinel Data Partner sites participating in this project had laboratory test results data approved for use in the MSDD. Thus, for the site that did not yet have laboratory test results data approved for use in the MSDD, the gold standard SUPREME-DM DataLink Algorithm and the SUPREME-DM DataLink Algorithm without laboratory test results identified the same number of patients.

The gold standard SUPREME-DM DataLink Algorithm (diagnoses, medications, and laboratory test results) identified 737,122 individuals in the MSDD. The SUPREME-DM algorithm without laboratory test results (diagnoses and medications only) identified 702,186 individuals, or approximately 5% fewer individuals as having DM from the other four sites with laboratory test results.

#### 4. Solberg Primary Any DM Algorithm

Case ascertainment with the primary Solberg algorithm for Any DM is detailed in Table 3.

**Table 3. Case Ascertainment of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard**

Primary Solberg Algorithm Applied to MSDD	Gold Standard (SUPREME-DM DataLink Algorithm)		Total
	Yes	No	
Yes	687,394 (A)	4,034 (B)	691,428
No	49,728 (C)		
Total	737,122		

##### a. SE of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites

The Solberg Primary Any DM Algorithm identified a total of  $687,394/737,122 = 93.3\%$  of the adults in the SUPREME-DM DataLink from the five sites as having Any DM. Across the five sites, the proportion identified as having Any DM ranged from 91.6% - 98.0%.

We considered all individuals who met the SUPREME-DM DataLink Algorithm Gold Standard (n= 737,122) as being true positive Any DM cases. Those who met both the Primary Solberg Any DM algorithm and the SUPREME-DM DataLink gold standard Algorithm (cell A) were considered the true positives identified in the Primary Solberg Any DM algorithm. The SE is the proportion of all positives that are true positives.

Discordant cell C (false negatives):

- 43,391/49,728 (87.3%) entered the gold standard by at least one lab criteria. The Solberg primary Any DM Algorithm does not include lab criteria.
- 5801/49,728 (11.7%) entered the gold standard by a metformin dispensing plus an outpatient diagnosis. The Solberg primary Any DM Algorithm excludes single-agent metformin.

##### b. SE of the Solberg Primary Any DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group

Age as of date of cohort entry:

20 – 44 Years:  $131,183/142,661 = 92.0\%$

45 – 64 Years:  $372,507/399,740 = 93.2\%$

65 – 74 Years:  $121,068/128,030 = 94.6\%$

≥ 75 Years:  $62,636/66,691 = 93.9\%$

##### c. SE of the Solberg Primary Any DM Algorithm versus SUPREME-DM Algorithm DataLink Gold Standard by Gender

Female:  $324,295/347,621 = 93.3\%$

Male:  $363,060/389,454 = 93.2\%$

Ambiguous:  $3/3 = 100\%$

Unknown:  $36/44 = 81.8\%$

**d. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites**

The Solberg Primary *Any* DM Algorithm identified a total 687,394/691,428= 99.4% of the adults in the SUPREME-DM DataLink as having *any* DM. Across the five sites, the proportion identified as having *any* DM ranged from 99.1% - 99.5%.

Discordant cell B: 4,006/4,034 (99.2%) entered using the Solberg Primary *Any* DM Algorithm by a single dispensing of an antidiabetic medication. In the gold standard, at least 2 dispensings of an antidiabetic medication or a combination of a medication dispensing with either a lab or an outpatient diagnosis is required.

**e. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Age Group**

20 – 44 Years: 131,183/132,482 = 99.0%  
 45 – 64 Years: 372,507/374,015 = 99.6%  
 65 – 74 Years: 121,068/121,634 = 99.5%  
 ≥ 75 Years: 62,636/63,297 = 99.0%

**f. PPV of the Solberg Primary *Any* DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Gender**

Female: 324,295/326,562 = 99.3%  
 Male: 363,060/364,827 = 99.5%  
 Ambiguous: 3/3 = 100%  
 Unknown: 36/36 = 100%

**5. Interpretation of the SE and PPV of the Solberg Primary *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard**

The Solberg Primary *Any* DM Algorithm does not include laboratory results criteria and most patients who are in the gold standard and not in the Solberg primary *Any* DM qualified by having met at least one laboratory criterion. ***This finding helps confirm the importance of laboratory results values for supplementing Any DM cohort identification.***

Most patients who qualified for the Solberg Primary *Any* DM Algorithm and did not meet the gold standard qualified through a single dispensing of an antidiabetic medication without meeting any other criteria for DM. These individuals could have DM but had either only a short enrollment period or had enrollment near the beginning or the end of the dataset timeframe (i.e., insufficient time to have a second qualifying DM parameter), or might not have DM. Without any second criterion being met, there is a lower degree of confidence that these individuals have DM. ***We recommend against a single medication dispensing criterion being sufficient to qualify as having Any DM.***

**6. Solberg Secondary *Any* DM Algorithm (Solberg Modified to include Emergency Department (ED) Visits in Diagnosis Criteria)**

Case ascertainment with the Solberg Secondary *Any* DM Algorithm is shown in Table 4.



**Table 4. Case Ascertainment with the Solberg Secondary Any DM Algorithm Modified to Include Emergency Department Visits versus the Gold Standard**

Solberg Secondary Algorithm Applied to MSDD	Gold Standard (SUPREME-DM DataLink Algorithm)		Total
	Yes	No	
Yes	689,726 (A)	13,669 (B)	703,395
No	47,396 (C)		
Total	737,122		

**a. SE of the Solberg Secondary Any DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites**

The Solberg Secondary Any DM Algorithm identified a total of 689,726/737,122 = 93.6% of the adults in the SUPREME-DM DataLink as having Any DM. Across the five sites, the proportion identified as having Any DM ranged from 91.9% - 98.1%.

Discordant cell C:

- 41,314/47,396 (87.2%) entered the gold standard by meeting at least one lab criteria. The Solberg Secondary Any DM Algorithm does not include lab criteria.
- 5,350/47,396 (11.2%) met the gold standard algorithm through a metformin fill + and one or more outpatient diagnosis code(s). The Solberg Secondary Any DM Algorithm excludes single-agent metformin.

**b. SE of the Solberg Secondary Any DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Age Group**

20 – 44 Years: 131,842/142,661 = 92.4%  
 45 – 64 Years: 373,801/399,740 = 93.5%  
 65 – 74 Years: 121,289/128,030 = 94.7%  
 ≥ 75 Years: 62,794/66,691 = 94.2%

**c. SE of the Solberg Secondary Any DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard by Gender**

Female: 325,334/347,621 = 93.6%  
 Male: 364,353/389,454 = 93.6%  
 Ambiguous: 3/3 = 100%  
 Unknown: 36/44 = 81.8%

**d. PPV of the Solberg Secondary Any DM Algorithm versus SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites**

The Solberg Secondary Any DM Algorithm identified a total of 689,726/ 703,396 = 98.1% of the adults in the SUPREME-DM DataLink as having Any DM. Across the five sites, the proportion identified as having Any DM ranged from 96.7% - 98.4%.

Discordant cell B:

- 9,843/13,669 (72.0%) qualified by having a diagnosis associated with an emergency department (ED) visit only. ED visits are not criteria in the gold standard.
- 3,647/13,669 (26.7%) qualified by having one antidiabetic medication dispensing only. In the gold standard, at least 2 dispensings of an antidiabetic medication or a combination of a medication dispensing with either a lab or an outpatient diagnosis is required.

#### e. PPV of the Solberg Secondary *Any* DM Algorithm versus Gold Standard by Age Group

20 – 44 Years: 131,842/135,815 = 97.1%

45 – 64 Years: 373,801/379,060 = 98.6%

65 – 74 Years: 121,289/123,201 = 98.4%

≥ 75 Years: 62,794/65,319 = 96.1%

#### f. PPV of the Solberg Secondary *Any* DM Algorithm versus Gold Standard by Gender

Female: 325,334/332,940 = 97.7%

Male: 364,353/370,416 = 98.4%

Ambiguous: 3/3 = 100%

Unknown: 36/36 = 100%

### 7. Interpretation of the SE and PPV of the Solberg Secondary *Any* DM Algorithm versus the Gold Standard

The Solberg Secondary *Any* DM Algorithm does not include glucose-related laboratory results criteria and most patients who are in met the gold standard algorithm and did not meet the Solberg Secondary *Any* DM Algorithm qualified by having met at least one laboratory criterion. ***This finding helps confirm the importance of glucose-related laboratory results values for supplementing Any DM cohort identification.***

Nearly 3/4 of the patients who qualified in the Solberg Secondary *Any* DM Algorithm who did not meet the gold standard algorithm qualified through a diagnosis made during an ED visit. Diagnosis during an ED visit is not a criterion for inclusion through the gold standard. Approximately 1/4 of the patients who qualified in the Solberg Secondary *Any* DM Algorithm and who did not meet the gold standard qualified through a single dispensing of an antidiabetic medication without meeting any other criteria for diabetes. In the gold standard, at least two dispensings of an antidiabetic medication are required or a medication dispensing in combination with a laboratory result or an outpatient diagnosis is required. Thus, these individuals who met the Solberg Secondary *Any* DM Algorithm criterion for a single antidiabetic medication dispensing could have DM either and had either only a short enrollment period or only had enrollment near the beginning or the end of the dataset timeframe (i.e., insufficient time to have a second qualifying DM parameter), or might not have DM. Without any second criterion being met, there is a lower degree of confidence that these individuals have DM. We recommend against a single medication dispensing criterion being sufficient to qualify as having DM.

### 8. Comparison of the Solberg Primary and Secondary *Any* DM Algorithms

SE of the two Solberg algorithms was similar (93.4% and 93.6%) relative to the gold standard. The Solberg Primary *Any* DM Algorithm had modestly better PPV than the Solberg Secondary *Any* DM Algorithm (99.4% vs. 98.1%) compared to the gold standard. Patients could meet the *Any* DM criteria in either of these Algorithms through a single dispensing of an antidiabetes medication. The SUPREME-DM DataLink Algorithm gold standard requires either two antidiabetes medication dispensings or a

medication dispensing in combination with a laboratory test result compatible with DM or a DM diagnosis code. Patients could be considered as having met the Solberg Secondary *Any* DM Algorithm criteria through a DM diagnosis code from an ED visit, whereas this was not a criterion in the Solberg Primary *Any* DM Algorithm.

## 9. Zgibor *Any* DM Algorithm

Case Ascertainment with Zgibor Algorithm for *Any* DM is shown in Table 5.

**Table 5. Case Ascertainment with the Zgibor Algorithm versus the Gold Standard**

Zgibor Algorithm Applied to MSDD	Gold Standard (SUPREME-DM DataLink Algorithm)		Total
	Yes	No	
Yes	703,332 (A)	3,895 (B)	707,227
No	33,790 (C)		
Total	737,122		

### a. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites

The Zgibor *Any* DM Algorithm identified a total of 703,332/737,122 = 95.4% of the adults in the SUPREME-DM DataLink as having *Any* DM. Across the five sites, the proportion identified as having *Any* DM ranged from 95.1% - 96.3%.

Discordant Cell C:

- 12,664/33,790 (37.5%) entered the gold standard by an inpatient diagnosis alone. The Zgibor *Any* DM Algorithm requires either two inpatient diagnoses or an inpatient diagnosis combined with a medication, outpatient diagnosis, ED diagnosis, or laboratory test.
- 12,525/33,790 (37.1%) entered the gold standard by at least one fasting glucose test. Fasting glucose is not a criterion in the Zgibor *Any* DM Algorithm.
- 3,737/33,790 (11.1%) enter the gold standard by at least two (at least one of which is non-metformin) medication dispensings. The Zgibor *Any* DM Algorithm requires at least one other criterion (diagnosis, lab) be met.
- 3,372/33,790 (10.0%) enter the gold standard by two elevated HbA1c test results. The Zgibor *Any* DM Algorithm requires at least one other criterion (diagnosis, medication, other laboratory test) be met.
- 649/33,790 (1.9%) enter the gold standard by two random glucose tests. The Zgibor *Any* DM Algorithm requires at least one other criterion (diagnosis, medication, other laboratory test) be met.

### b. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group

20 – 44 Years: 136,865/142,661 = 95.9%

45 – 64 Years: 384,769/399,740 = 96.3%

65 – 74 Years: 121,981/128,030 = 95.3%

≥ 75 Years: 59,717/66,691 = 89.5%

**c. SE of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Gender**

Female: 331,076/347,621 = 95.2%

Male: 372,211/389,454 = 95.6%

Ambiguous: 3/3 = 100%

Unknown: 42/44 = 95.5%

**d. PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard across Sites Combined and by Individual Sites**

The Zgibor *Any* DM Algorithm identified a total of 703,332/707,227 = 99.4% of the adults in the SUPREME-DM DataLink as having *Any* DM. Across the five sites, the proportion identified as having *Any* DM ranged from 99.1% - 99.6%.

Discordant Cell B:

- 1,477/3,895 (37.9%) entered in the Zgibor *Any* DM Algorithm by a single outpatient diagnosis. In the gold standard, either two outpatient diagnoses or an outpatient diagnosis in combination with a qualifying laboratory result or a medication dispensing is required.
- 1451/3895 (37.3%) enter in the Zgibor *Any* DM Algorithm by the combination of an ED visit diagnosis and an outpatient diagnosis. ED visits are not criteria in the gold standard.
- 403/3895 (10.3%) enter in the Zgibor *Any* DM Algorithm by a combination of an ED visit diagnosis + an HbA1c test, random glucose test, or medication. ED visits are not criteria in the gold standard.

**e. PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Age Group**

20 – 44 Years: 136,865/137,968 = 99.2%

45 – 64 Years: 384,769/386,362 = 99.6%

65 – 74 Years: 121,981/122,557 = 99.5%

≥ 75 Years: 59,717/60,340 = 99.0%

**f. PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard by Gender**

Female: 331,076/333,113 = 99.4%

Male: 372,211/374,069 = 99.5%

Ambiguous: 3/3 = 100%

Unknown: 42/42 = 100%

**10. Interpretation of the SE and PPV of the Zgibor *Any* DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard**

Among patients who are considered as having *Any* DM based on the gold standard, but who did not meet criteria for *Any* DM using the Zgibor *Any* DM Algorithm, over 1/3 entered the SUPREME-DM DataLink gold standard by an inpatient diagnosis alone. One inpatient diagnosis is not sufficient in the Zgibor *Any* DM Algorithm. About another 1/3 of patients who are in the gold standard but did not meet criteria using the Zgibor *Any* DM Algorithm entered the gold standard by fasting glucose tests. Fasting glucose is not a criterion in the Zgibor *Any* DM Algorithm. Most of the other patients in the gold

standard who did not meet criteria using the Zgibor *Any* DM Algorithm met gold standard criteria by meeting the same criterion a second time on a different date, whereas the Zgibor *Any* DM Algorithm required a separate criterion be met (does not include outpatient diagnosis) and these patients did not meet a second criterion.

Qualifying for the gold standard *Any* DM algorithm based on an outpatient diagnosis requires either two outpatient diagnoses or an outpatient diagnosis in combination with a qualifying laboratory result or a medication dispensing. Over 1/3 of patients who were considered to have *Any* DM based on the Zgibor *Any* DM Algorithm who did not meet the gold standard had a single outpatient diagnosis. Nearly one-half of the other patients who qualified in the Zgibor *Any* DM Algorithm who did not meet the gold standard algorithm qualified through the combination of an ED visit diagnosis with an outpatient diagnosis with or without a laboratory test result or medication dispensing. ED visits are not criteria in the SUPREME-DM DataLink Algorithm gold standard.

***If the goal is to maximize SE, then the Zgibor approach to allow only one outpatient diagnosis to qualify an individual as having Any DM could be considered. However, the workgroup does not recommend this because it results in including more falsely positive patients: previous studies by O'Connor et al<sup>8</sup> and by Solberg and colleagues<sup>1</sup> demonstrated that using only one DM diagnostic code identified many patients who did not have DM, providing an unacceptably low PPV.***

## **11. Comparison of the Solberg Primary *Any* DM Algorithm, the Solberg Secondary *Any* DM Algorithm, and the Zgibor *Any* DM Algorithm versus the Gold Standard**

The SE of the Zgibor *Any* DM Algorithm was modestly better than that of either of the Solberg *any* DM Algorithms (SE of the Zgibor *Any* DM Algorithm, 95.4% vs. 93.4% and 93.6%, respectively for the Solberg Primary and Solberg Secondary *Any* DM Algorithms) in reference to the gold standard. The PPV of the Zgibor *Any* DM Algorithm was similar to that of the Solberg Primary *Any* DM Algorithm and modestly better than that of the Solberg Secondary *Any* DM Algorithm (PPV of the Zgibor *Any* DM Algorithm, 99.2% vs. 99.4% vs. 98.1%, respectively for the Solberg Primary and Secondary *Any* DM Algorithms) compared to the gold standard. The differences in SE and PPV of these algorithms relative to the gold standard were largely due to considering a single outpatient diagnosis as sufficient to qualify as having *any* DM (the Zgibor *Any* DM Algorithm), including diagnoses during ED visits (the Solberg Primary *Any* DM Algorithm and the Zgibor *Any* DM Algorithm), and including laboratory tests as criteria (Zgibor *Any* DM Algorithm). Among these three *Any* DM Algorithms, the Zgibor Algorithm had both the best SE and PPV.

### **C. COMPARISON OF PUBLISHED ALGORITHMS FOR PATIENTS WITH T1DM IN THE MSDD**

Only the two Klompas<sup>4</sup> T1DM algorithms met the workgroup requirement that the algorithm must have been tested in adults (Table 6). We tested both the Klompas Primary Algorithm and the Klompas Optimized Algorithm. However, because neither of the Klompas algorithms had been externally validated, neither was considered a gold standard. We therefore could only compare the number of cases identified using the two Klompas algorithms to each other. The Klompas Optimized T1DM Algorithm originally included C-peptide or DM autoantibodies (Table 6, Pathways 4 and 5), but these laboratory test results are not available in the SUPREME-DM DataLink or the MSDD. The Klompas Optimized T1DM Algorithm initially tested therefore only included Pathways 1 - 3.

We required individuals to meet the gold standard *Any* DM Definition (the SUPREME-DM DataLink definition) before applying either T1DM algorithm. We did this because the T1DM algorithms were not designed to identify individuals with *Any* DM, but rather to determine who has T1DM in a population

with Any DM. In this way the workgroup approach was consistent with the methodology used by Klompas et al in deriving the T1DM algorithms.<sup>4</sup>

**Table 6. Algorithms for Type 1 Diabetes Mellitus**

Algorithm Identifier	Original Algorithm	Tested/Modified Algorithm	Notes from 15 Cohorts WG <sup>2</sup>
<p><u>1. Primary Type 1 Diabetes Algorithm</u></p> <p>Diagnosis codes and medications</p> <p>Klompas<sup>4</sup></p>	<p>1) a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes &gt;0.5], AND</p> <p>2) no prescription for a non-insulin antidiabetic drug (excluding metformin)]</p> <p>Restating #2 for clarity: A prescription for insulin or metformin is allowed, but a prescription for any other antidiabetes medication results in exclusion</p>	Original algorithm	Workgroup recommendation for the primary Type 1 Diabetes algorithm.
<p><u>2. Optimized Type 1 Diabetes Algorithm</u></p> <p>Diagnosis codes and medications</p> <p>Klompas Optimized<sup>4</sup></p>	<p>Pathway 1:</p> <p>1) a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes &gt;0.5, AND</p> <p>2) no prescription for a non-insulin antidiabetic drug (excluding metformin)]</p> <p>Restating #2 for clarity: A prescription for insulin or metformin is allowed, but a prescription for any other antidiabetes medication results in exclusion</p> <p>OR Pathway 2:</p> <p>1) a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes &gt;0.5, AND</p> <p>2) a prescription for glucagon</p> <p>OR Pathway 3:</p> <p>Prescription for urine acetone test strips</p> <p>OR Pathway 4:</p> <p>C-peptide negative</p> <p>OR Pathway 5:</p> <p>diabetes autoantibodies positive</p>	<p>Pathway 1 (no change):</p> <p>1) a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes &gt;0.5, AND</p> <p>2) no prescription for a non-insulin antidiabetic drug (excluding metformin)]</p> <p>Restating #2 for clarity: A prescription for insulin or metformin is allowed, but a prescription for any other antidiabetes medication results in exclusion</p> <p>OR Pathway 2 (no change):</p> <p>1) a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes &gt;0.5, AND</p> <p>2) a prescription for glucagon</p> <p>OR Pathway 3: Prescription for urine acetone test strips (No change as initially tested. Workgroup does not recommend its use as a criterion; see text below)</p> <p>Pathways 4 and 5: Removed because C-peptide and diabetes autoantibodies are not available</p>	<p>Klompas et al's optimized algorithm; use if requisite laboratory data become widely available within the distributed database, if not, drop the definition components requiring C-peptide and diabetes autoantibodies, as PPV and SE are still high for such a definition</p>

## 1. Klompas Primary T1DM Algorithm

The number of patients identified with T1DM using the Klompas Primary T1DM algorithm is shown in Table 7.

**Table 7. Patients Identified with T1DM in the MSDD Based on the Klompas Primary T1DM Algorithm**

Klompas Primary T1DM Algorithm Applied to MSDD	Patients in the MSDD Previously Identified using the Gold Standard SUPREME-DM DataLink Algorithm as having Any Diabetes		Total
	Yes	No	
Yes	24,850 (A)	1,780 (B)	
No	712,272 (C)		
Total	737,122		

### a. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm across Sites Combined and by Individual Sites

The Klompas Primary Algorithm identified a total of 24,850 (3.4%) of the adults with Any DM in the MSDD from these five sites as having T1DM. Across the five sites, the proportion identified as having T1DM ranged from 2.2% - 6.8%.

The 1,780 individuals in shaded cell B in Table 7 did not meet the gold standard Any DM Definition (the SUPREME-DM DataLink Algorithm), even though they were identified by the Klompas Primary Algorithm. Most (1,418 or 79.7%) of these 1,780 cases had just a single T1DM diagnosis which may or may not have been an outpatient diagnosis. The remaining 362 (20%) cases entered with two or more diagnoses that were either not outpatient diagnoses or did not fall on different days within two years (as required by the SUPREME-DM DataLink Algorithm).

### b. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm by Age Group

20 – 44 Years: 15,192/142,661 = 10.6%

45 – 64 Years: 8,076/399,740 = 2.0%

65 – 74 Years: 1,061/128,030 = 0.8%

≥ 75 Years: 521/66,691 = 0.8%

The denominator for each age group is the number of individuals with Any DM in that age group.

### c. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Primary Algorithm by Gender

Female: 11,536/347,621 = 3.3%

Male: 13,314/389,454 = 3.4%

Ambiguous: 0/3 = 0%

Unknown: 0/44 = 0%

The denominator for each gender is the number of individuals with Any DM with that gender.



## 2. Klompas Optimized T1DM Algorithm

The number of patients identified with T1DM using the Klompas Optimized Algorithm is in Table 8.

**Table 8. Patients Identified with T1DM in the MSDD Based on the Klompas Optimized Algorithm**

Klompas Optimized Algorithm Applied to MSDD	Patients in the MSDD Previously Identified using the Gold Standard SUPREME-DM DataLink Algorithm as having <i>Any</i> Diabetes		Total
	Yes	No	
Yes	26,418 (A)	2,189 (B)	
No	710,704 (C)		
Total	737,122		

### a. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm across Sites Combined and by Individual Sites

The Klompas Optimized T1DM Algorithm (Pathways 1 – 3) identified 26,418 (3.6%) of the adults with *Any* DM in the MSDD at these five sites as having T1DM. Across the five sites, the proportion identified as having T1DM ranged from 2.4% - 7.4%.

The 2,189 individuals in shaded cell B in Table 8 did not meet the gold standard *Any* DM Definition (the SUPREME-DM DataLink Algorithm), even though they were identified by the Klompas Optimized Algorithm. Most (1,780; 81.3%) of these 2,189 cases are the same cases as previously discussed for the Klompas Primary Algorithm. The remaining 409 (18.7%) cases met the Klompas Optimized Algorithm criteria by having a prescription for urine acetone test strips. Because neither the SUPREME-DM DataLink nor the MSDD has C-peptide or diabetes autoantibody laboratory test results, no cases could enter through those pathways (Pathways 4 and 5) when this algorithm was initially tested.

### b. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm by Age Group

20 – 44 Years:  $15,926/142,661 = 11.2\%$

45 – 64 Years:  $8,745/399,740 = 2.2\%$

65 – 74 Years:  $1,189/128,030 = 0.9\%$

≥ 75 Years:  $558/66,691 = 0.8\%$

The denominator for each age group is the number of individuals with *Any* DM in that age group.

A higher proportion of people with T1DM were identified in younger age groups. It was expected that T1DM prevalence would be higher in the young.

### c. Numbers and Proportions of Patients in MSDD Identified as Having T1DM by the Klompas Optimized Algorithm by Gender

Female:  $12,469/347,621 = 3.6\%$

Male:  $13,949/389,454 = 3.6\%$

Ambiguous:  $0/3 = 0\%$

Unknown:  $0/44 = 0\%$

The denominator for each gender is the number of individuals with *Any* DM with that gender.



### 3. Comparison of Klompas Primary T1DM and Klompas Optimized T1DM Algorithm with Neither Considered a Gold Standard

In Table 9 the two Klompas T1DM Algorithms are compared. The T1DM cases identified using the Klompas Primary Algorithm is a subset of the cases identified using the Klompas Optimized Algorithm (Pathways 1 – 3). That is, all T1DM cases identified by the Klompas Primary Algorithm are also identified by the Klompas Optimized Algorithm and the Klompas Optimized Algorithm identified additional cases with T1DM.

**Table 9. Comparisons of the Klompas Primary T1DM Algorithm and the Klompas Optimized T1DM Algorithm with Neither Considered a Gold Standard**

Klompas Primary T1DM Algorithm	Klompas Optimized T1DM Algorithm		Total
	Yes	No	
Yes	24,850 (A)	0 (B)	24,850
No	1,568 (C)	710,704 (D)	712,272
Total	26,418	710,704	737,122

The Klompas Optimized Algorithm identified 1,568 (6.7%; range across sites 2.1% –9.8%) more cases as T1DM than the Klompas Primary Algorithm. These additional cases identified by the Klompas Optimized Algorithm were due to the following Pathways:

- 309 (19.7%) entered by Pathway 2 only (diagnosis + glucagon)
- 1,086 (69.3%) entered by Pathway 3 only (urine acetone test strips)
- 173 (11.0%) entered by Pathways 2 and 3 (diagnosis + glucagon + urine acetone test strips)

There were some differences by age in the additional percentage of cases with T1DM identified by the Klompas Optimized Algorithm compared to the Klompas Primary Algorithm:

20 – 44 Years: 734 additional cases = 4.8% (734/15,192)  
 45 – 64 Years: 669 additional cases = 8.3 % (669/8,076)  
 65 – 74 Years: 128 additional cases = 12.1% (128/1,061)  
 ≥ 75 Years: 37 additional cases = 7.1% (37/521)

The denominator for each age group is the number of T1DM cases in that age group identified by the Klompas Optimized Algorithm.

The Klompas Optimized Algorithm identified 8.1% (n = 933) more females and 4.8% (n=635) more males with T1DM than the Klompas Primary Algorithm.

### 4. Interpretation of Case Identification using the Klompas Primary and Optimized T1DM Algorithms

The version of the Klompas Optimized Algorithm we tested is somewhat preferred over the Klompas Primary Algorithm for identifying patients with T1DM in the MSDD. However, in the tested Klompas Optimized Algorithm, 1,086 patients were classified as T1DM on the basis of Pathway 3, a urine acetone test strip dispensing *alone*. Because we do not consider a urine acetone test strip dispensing by itself is a sufficiently robust criterion to consider an individual as having T1DM, at this time the workgroup cannot recommend including Pathway 3 in the T1DM algorithm (considering urine acetone test strip dispensing alone as a criterion for classifying as T1DM alone would first need to be validated through chart review).

Omitting the individuals identified as having T1DM through Pathway 3 only, the individuals from these five sites in the MSDD identified as having T1DM decreases from 26,418 (3.6%) to 25,332 (3.4%). Also, because laboratory test results for C-peptide and diabetes autoantibodies are not available in the MSDD, in the Klompas Optimized Algorithm initially tested, we did not include Pathways 4 and 5 (C-peptide and diabetes autoantibodies laboratory test results). We recommend that only the Klompas Optimized Algorithm Pathways 1 and 2 be included as T1DM criteria if applying the Klompas Optimized Algorithm to the MSDD.

Approximately 5% of the adult American population with diagnosed diabetes is believed to have T1DM.<sup>42-44</sup> The prevalence of T1DM in American youth is fairly well defined using data from the SEARCH for Diabetes in Youth study.<sup>36,43,45</sup> However, similar data does not exist for adults. The National Health And Nutrition Examination Survey (NHANES) collects information such as age of diabetes diagnosis, current insulin use, and age of insulin initiation, but not on diabetes type.<sup>5</sup> Researchers have thus used the following (or similar) definitions when attempting to identify T1DM using NHANES data: 1) started insulin with 1 year of diabetes diagnosis, and 2) currently using insulin, and 3) diagnosed with diabetes under age 30 or 40. Using these definitions, the proportion of individuals with T1DM diabetes (both youth and adults) based on NHANES 1999-2010 data is 4.6% (age 30 cut point) to 4.8% (age 40 cut point). However, this definition clearly misses individuals in whom T1DM develops after age 40. In addition, the SE for the optimized T1DM algorithm in the Klompas article was 65% in their validation data set.<sup>4</sup> There is thus considerable uncertainty in T1DM case identification using the Klompas algorithms, but it is most likely that the 3.4% of the MSDD population we identified with T1DM using the Klompas Optimized Algorithm Pathways 1 and 2 is an under-ascertainment of cases.

#### **D. CASE ASCERTAINMENT USING PUBLISHED ALGORITHMS FOR PATIENTS WITH T2DM IN THE MSDD**

We tested the two eMERGE “case” T2DM algorithms.<sup>6</sup> The eMERGE Primary T2DM Algorithm includes diagnosis codes and medications (Table 10). The eMERGE Secondary T2DM Algorithm includes diagnosis codes, medications, and laboratory test results (Table 10). Comparing the eMERGE Primary and Secondary Algorithms at the sites with laboratory results in the MSDD allowed us to estimate the impact of unavailable laboratory results data on these algorithms.

Although the two eMERGE T2DM algorithms begin with the entire universe of individuals as the starting point (i.e., not limited to individuals with *Any DM*), we required individuals to have met the gold standard *Any DM Definition* before applying the T2DM algorithms (i.e., we started with the individuals in the MSDD identified as having *Any DM*). In this way the workgroup approach differed from the methodology used by the eMERGE team.

The workgroup could not confirm that the eMERGE Primary and Secondary Algorithms had been externally validated. Neither eMERGE T2DM Algorithm was considered a gold standard.

**Table 10. Algorithms for Type 2 Diabetes Mellitus**

Algorithm Identifier	Original Algorithm	Tested/Modified Algorithm	Notes from 15 Cohorts WG <sup>2</sup>
<p><u>8. Primary Type 2 Diabetes Algorithm</u></p> <p>Diagnosis codes and medications</p> <p>eMERGE “case”<sup>6</sup> algorithm without labs</p>	<p>Pathway 1</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND</li> <li>2) ≥ 1 type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND</li> <li>3) Any type 2 diabetes medication, AND</li> <li>4) Any type 1 diabetes medication (insulin, pramlintide), AND</li> <li>5) Date of type 2 diabetes medication &lt; date of type 1 diabetes medication</li> </ol> <p>OR Pathway 2</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND</li> <li>2) ≥ 1 type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND</li> <li>3) Any type 2 diabetes medication, (in this case, we are allowing metformin due to “AND” criterion) AND</li> <li>4) No type 1 diabetes medication (insulin, pramlintide)</li> </ol> <p>OR Pathway 5</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND</li> <li>2) ≥ 1 type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND</li> <li>3) Any type 1 diabetes medication (insulin, pramlintide), AND</li> <li>4) No type 2 diabetes medication, AND</li> <li>5) ≥ 2 Type 2 diabetes diagnosis by physician</li> </ol>	<p>Only Pathways 1 and 2.</p> <p>We are not able to define Pathway 5 using MSCDM (due to the “AND ≥ 2 Type 2 diabetes diagnosis by physician,” because it cannot be confirmed that the diabetes diagnosis was assigned by a physician, so that pathway was not included</p>	<p>Workgroup had no recommendation for T2DM algorithm</p>
<p><u>9. Secondary Type 2 Diabetes Algorithm</u></p> <p>Diagnosis codes, medications, and labs</p> <p>eMERGE “case”<sup>6</sup> algorithm with labs</p>	<p>Pathway 1</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND</li> <li>2) ≥ 1 type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND</li> <li>3) Any type 2 diabetes medication (including metformin), AND</li> <li>4) Any type 1 diabetes medication (insulin, pramlintide), AND</li> <li>5) Date of type 2 diabetes medication &lt; date of type 1 diabetes medication</li> </ol> <p>OR Pathway 2</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND</li> <li>2) ≥ 1 type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND</li> <li>3) Any type 2 diabetes medication (including metformin), AND</li> <li>4) No type 1 diabetes medication (insulin, pramlintide)</li> </ol> <p>OR Pathway 3</p> <ol style="list-style-type: none"> <li>1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3),</li> </ol>	<p>Pathways 1 through 4.</p> <p>The SUPREME-DM DataLink algorithm applies fasting glucose ≥ 126 mg/dl and a random glucose ≥ 200 mg/dl in keeping with the 2011 and 2015 ADA DM diagnosis criteria. The eMERGE criteria apply fasting glucose ≥ 125 mg/dl and a random glucose &gt; 200 mg/dl. We modified the eMERGE algorithm Pathways 3 and 4 to apply fasting glucose ≥ 126 mg/dl and</p>	<p>Workgroup had no recommendation for T2DM algorithm</p>

Algorithm Identifier	Original Algorithm	Tested/Modified Algorithm	Notes from 15 Cohorts WG <sup>2</sup>
	regardless of source, AND 2) $\geq 1$ type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless of source, AND 3) No type 1 diabetes medication (insulin, pramlintide), AND 4) No type 2 diabetes medication (including metformin), AND 5) Random plasma glucose $> 200$ mg/dL OR Fasting plasma glucose $\geq 125$ mg/dL OR HgA1c $\geq 6.5\%$  OR Pathway 4 1) No type 1 diabetes ICD-9 codes (250.x1, 250.x3), regardless of source, AND 2) No type 2 diabetes ICD-9 codes (250.x0, 250.x2; excluding 250.10, 250.12), regardless or source, AND 3) Any type 2 diabetes medication (metformin), AND 4) Random plasma glucose $> 200$ mg/dL OR Fasting plasma glucose $\geq 125$ mg/dL OR HgA1c $\geq 6.5\%$  OR Pathway 5 (See Primary Type 2 Diabetes Algorithm and next column)	random glucose $\geq 200$ mg/dl.  We are not able to define Pathway 5 using MSCDM (due to the “AND $\geq 2$ Type 2 diabetes diagnosis by physician,” because it cannot be confirmed that the diabetes diagnosis was assigned by a physician, so that pathway was not included	

### 1. eMERGE Primary T2DM Algorithm

The number of patients identified with T2DM using the eMERGE Primary Algorithm is shown in Table 11.

**Table 11. Patients Identified with T2DM in the MSDD Based on the eMERGE Primary Algorithm**

eMERGE Primary T2DM Algorithm Applied to MSDD	Patients in the MSDD Previously Identified using the Gold Standard SUPREME-DM DataLink Algorithm as having Any Diabetes		Total
	Yes	No	
Yes	432,912 (A)	1,359 (B)	
No	304,210 (C)		
Total	737,122		

#### a. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Primary Algorithm across Sites Combined and by Individual Site

The eMERGE Primary Algorithm identified 432,912 (58.7%) of the adults in the MSDD from the five participating sites as having T2DM. Across the five sites, the proportion identified with T2DM ranged from 49.6%-61.5%.

If we had not limited our observations to individuals in the MSDD first identified with Any DM using the gold standard SUPREME-DM DataLink criteria, 1,359 additional individuals (cell B) in the MSDD would have met T2DM criteria using the eMERGE Primary Algorithm. Nearly all (1,354; 99.6%) of these additional cases would have qualified through Pathway 2. The remaining 5 (0.4%) cases qualified through Pathway 1. Criteria would have been met because the eMERGE Primary T2DM Algorithm includes visits other than, or in addition to, outpatient or inpatient visits, only requires one diagnosis

regardless of source, and the diagnosis and dispensing were not required to be within 730 days of each other (required in the gold standard).

**b. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Primary Algorithm by Age Group**

20 – 44 Years: 86,410/142,661 = 60.6%  
 45 – 64 Years: 251,910/399,740 = 63.0  
 65 – 74 Years: 68,455/128,030 = 53.5%  
 > 75 Years: 26,137/66,691 = 39.2%

The denominator for each age group is the number of individuals with *Any* DM in that age group.

**c. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by eMERGE Primary Algorithm by Gender**

Female: 200,434/347,621 = 57.7%  
 Male: 232,448/389,454 = 59.7%  
 Ambiguous: 2/3 = 66.7%  
 Unknown: 28/44 = 63.6%

The denominator for each gender is the number of individuals with *Any* DM with that gender.

**2. eMERGE Secondary T2DM Algorithm**

The number of patients identified with T2DM from the eMERGE Secondary Algorithm is shown in Table 12.

**Table 12. Patients Identified with T2DM in the MSDD Based on the eMERGE Secondary Algorithm**

eMERGE Secondary T2DM Algorithm Applied to MSDD	Patients in the MSDD Previously Identified using the Gold Standard Algorithm as having <i>Any</i> Diabetes		Total
	Yes	No	
Yes	521,450 (A)	4,293 (B)	
No	215,672 (C)		
<b>Total</b>	<b>737,122</b>		

**a. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm across Sites Combined and by Individual Site**

The eMERGE Secondary Algorithm identified 521,450 (70.7%) of the adults in the MSDD from these five participating sites as having T2DM. Across the five sites, the proportion identified as having T2DM ranged from 49.6%-76.1%.

If we had not limited identification to individuals in the MSDD first identified with *Any* DM using the gold standard SUPREME-DM DataLink criteria, an additional 4,293 individuals (cell B) in the MSDD would have been included when applying the eMERGE Secondary Algorithm. These include the 1,359 individuals identified using the eMERGE Primary Algorithm (based on the same criteria as in the eMERGE Primary Algorithm), as well as 2,934 additional cases. Among the additional 2,934 cases, 2,698 entered through Pathway 3 and 236 entered through Pathway 4. The reasons for meeting the eMERGE Secondary Algorithm criteria and not the gold standard criteria again include that visits other than

outpatient or inpatient visits are considered for diagnosis for the eMERGE Secondary Algorithm. Also, diagnosis, dispensing, and/or laboratory criteria are not required to occur within 730 days of each other (which is required in the gold standard), and only one diagnosis was required regardless of source.

**b. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm by Age Group**

20 – 44 Years: 96,631/142,661 = 67.7%  
 45 – 64 Years: 294,998/399,740 = 73.8%  
 65 – 74 Years: 89,794/128,030 = 70.1%  
 ≥ 75 Years: 40,027/66,691 = 60.0%

The denominator for each age group is the number of individuals with *Any* DM in that age group.

**c. Numbers and Proportions of Patients in MSDD Identified as Having T2DM by the eMERGE Secondary Algorithm by Gender**

Female: 243,025/347,621 = 69.9%  
 Male: 278,387/389,454 = 71.5%  
 Ambiguous: 3/3 = 100%  
 Unknown: 35/44 = 79.5%

The denominator for each gender is the number of individuals with *Any* DM with that gender.

**3. Comparison of the eMERGE Primary T2DM Algorithm and the eMERGE Secondary T2DM Algorithm with Neither Considered a Gold Standard**

The two eMERGE T2DM algorithms are compared in Table 13. The eMERGE Primary Algorithm is a subset of the eMERGE Secondary Algorithm. That is, all T2DM cases identified by the eMERGE Primary Algorithm are also identified by the eMERGE Secondary Algorithm and the eMERGE Secondary Algorithm identifies additional cases as T2DM.

**Table 13. Comparisons of T2DM eMERGE Primary Algorithm (without Laboratory Test Results) and eMERGE Secondary Algorithm (with Laboratory Test Results) with Neither Considered a Gold Standard**

eMERGE Primary Algorithm	eMERGE Secondary Algorithm		Total
	Yes	No	
Yes	432,912 (A)	0 (B)	432,912
No	88,538 (C)	215,672 (D)	304,210
<b>Total</b>	521,450	215,672	737,122

The eMERGE Secondary Algorithm identified 88,538 (20.5%; range across sites 0% [the site without laboratory results data] to 25.1%) more cases as T2DM than the eMERGE Primary Algorithm. These additional cases identified by the eMERGE Secondary Algorithm were due to the following:

- 84,866 (95.9% of 88,538) entered by T2DM diagnosis and random or fasting glucose or HbA1c (Pathway 3)
- 3,672 (4.1% of 88,538) entered by T2DM diagnosis, T2DM medication, and random or fasting glucose or HbA1c (Pathway 4)

There were differences by age in the additional cases with T2DM identified by the eMERGE Secondary Algorithm:

20 – 44 Years: 10,221 additional cases = 11.8% (10,221/86,410)  
 45 – 64 Years: 43,088 additional cases = 17.1% (43,088/251,910)  
 65 – 74 Years: 21,339 additional cases = 31.2% (21,339/68,455)  
 ≥ 75 Years: 13,890 additional cases = 53.1% (13,890/26,137)

The denominator for each age group is the number of individuals with T2DM identified by the eMERGE Secondary Algorithm in that age group. The eMERGE Secondary Algorithm identified an important number and proportion of additional cases among individuals ages 65 years and older.

The eMERGE Secondary Algorithm identified 21.2% (n = 42,591) more females and 19.8% (n=45,939) more males with T2DM than the eMERGE Primary Algorithm.

#### 4. Interpretation of Case Identification using the T2DM eMERGE Primary and Secondary Algorithms

The eMERGE Secondary Algorithm is somewhat preferred over the eMERGE Primary Algorithm for identifying patients T2DM in the MSDD. If an eMERGE T2DM algorithm is to be employed, the eMERGE Secondary Algorithm, including Pathways 1 through 4 should be used (modified to be consistent with the 2011 and 2015 ADA DM diagnosis criteria of fasting glucose  $\geq$  126 mg/dl and random glucose  $\geq$  200 mg/dl).

In the MSDD population, employing the eMERGE Secondary Algorithm identified 70.7% of adults as having T2DM, compared to 58.7% of adults identified with T2DM employing the eMERGE Primary Algorithm. However, the proportion of cases in the MSDD identified with T2DM varied widely across the participating sites (49.6%-76.1%), and was lower than expected at all sites.

The vast majority of additional cases identified using the eMERGE Secondary Algorithm were from combining T2DM diagnosis codes with a random or fasting glucose or HbA1c. That is, most of the additional cases identified using the eMERGE Secondary Algorithm are T2DM cases from Data Partner sites that have laboratory results data in the MSDD. The eMERGE Primary and eMERGE Secondary Algorithms identified a similar number of cases from the participating Data Partner site that does not have laboratory results data. Under-ascertainment of T2DM will occur at Data Partner sites without laboratory results data in the MSDD. ***Use of glucose-related laboratory test result values criteria increases the number of Any DM cases and the number of T2DM cases identified.***

Importantly, compared to the eMERGE Primary Algorithm, the eMERGE Secondary Algorithm identified a substantial number of additional T2DM cases among individuals' aged 65 years and older (ages 65 – 74: eMERGE Primary Algorithm = 68,455 of 128,030 [53.5%] versus eMERGE Secondary Algorithm = 89,794 of 128,030 [70.1%]; ages  $\geq$  75: eMERGE Primary Algorithm = 26,137 of 66,691 [39.2%] versus eMERGE Secondary Algorithm = 40,027 of 66,691 [60.0%]).

#### E. PATIENTS WITH ANY DM WHO HAVE DM OF UNCERTAIN TYPE AFTER APPLYING T1DM AND T2DM ALGORITHMS

Fully 25.7% (189,593 of 737,122; range across the five sites: 21.5% to 43.1%) of individuals within the MSDD identified from the participating sites as having Any DM did not meet either the criteria to be classified as having T1DM based on the Klompas Optimized T1DM Algorithm or as having T2DM based on the eMERGE Secondary T2DM Algorithm definitions. Examining these individuals informs why they



do not meet either T1DM or T2DM criteria, potentially aiding in refining the algorithms for use within the MSDD.

### 1. Why Individuals Did Not Meet eMERGE T2DM Algorithm Criteria

Exploring why individuals did not meet criteria in the eMERGE T2DM Secondary Algorithm reveals that many of those individuals only partially met criteria in Pathways 1 - 4. For example:

- Pathway 1: 61,949 of 189,593 (32.7%) met the two diagnosis criteria (no T1DM ICD-9 codes and > 1 T2MD ICD-9 codes), but did not meet the medication and date order criteria (any T2DM medication and any T1DM medication and date of T2DM medication < date of T1DM medication)
- Pathway 1: 31,866 (16.8%) met both diagnosis and medication criteria, but the date order criterion was not met
- Pathway 2: 27,802 (14.7%) met the T2DM diagnosis criteria (> 1 T2DM diagnosis) and T2DM medication criterion (any T2DM medication), but also had some T1DM diagnosis codes and medications (exclusions in Pathway 2)
- Pathway 3: 61,949 (32.7%) met the first four criteria (no T1DM diagnosis, > 1 T2DM diagnosis, no T1DM or T2DM medications) but did not meet the laboratory criteria (random plasma glucose > 200 mg/dL, fasting plasma glucose > 126 mg/dl, or HgA1c > 6.5%)
- Pathway 4: 68,272 (36.0%) had T2DM codes (exclusion in Pathway 4), no type 2 medications, and no qualifying laboratory result values

The eMERGE algorithm does not include the option of being defined as having T2DM using only  $\geq 2$  T2DM diagnosis codes AND no T1DM diagnosis codes (with or without medications and with or without laboratory results available). We therefore evaluated what proportion of the 189,593 patients would have been considered as meeting T2DM criteria based on  $\geq 2$  T2DM diagnosis codes AND no T1DM diagnosis codes. Fully 48.7% of the individuals who did not meet eMERGE T2DM Algorithm criteria would have met criteria for having T2DM based on these diagnosis codes criteria. Importantly, the range across sites was from 42.1% to 63.2% of the uncertain DM patients who would be reclassified into the T2DM group. The highest proportion reclassified was at the site that did not have laboratory results available in the MSDD.

It is likely that the majority of individuals who did not meet criteria required in the eMERGE T2DM Secondary Algorithm Pathways (or the T1DM algorithm) actually have T2DM, but a) were treated with an oral medication and insulin started at the same time, b) had predominantly T2DM diagnosis codes, but also had one or more T1DM diagnosis codes, c) were not treated with an antidiabetic medication (whether not prescribed or never dispensed), or d) had other diabetes medication use patterns such as profound non-adherence. Exploration of these important subgroups of patients (particularly those patients who do not meet the 2 T2DM diagnosis codes AND no T1DM diagnosis codes modification to the criteria for T2DM) will be important to determine whether they do appear to have T2DM, if they have T1DM, if they have a condition such as latent autoimmune diabetes of adults (LADA; often supported by low C-peptide and diabetes autoantibodies rather than insulin resistance), or if they have other uncommon types of DM that are not well-coded using diabetes diagnosis codes (e.g., monogenetic defects in beta-cell function such as maturity-onset diabetes of the young or MODY, drug-induced diabetes such as due to glucocorticoids, or diseases affecting the pancreas such as cystic fibrosis).<sup>46</sup>



Comparing the proportion of individuals identified using the eMERGE Primary (no labs) and Secondary (with labs) Algorithms between sites with and without laboratory results in the MSDD helps inform us about patients who likely have T2DM that are not identified when laboratory results for glucose or HbA1c tests are not available in the MSDD. As expected, the participating site that does not have laboratory test results in the MSDD identified 49.6% of its patients with *Any DM* in the MSDD as having T2DM using both the eMERGE Primary and Secondary T2DM Algorithms (Table 14). In contrast, the four participating sites with laboratory test results in the MSDD identified 53.3% to 61.5% of their patients with *Any DM* in the MSDD as having T2DM using the eMERGE Primary Algorithm (without labs) and increased their proportion of cases identified using the eMERGE Secondary T2DM Algorithm (with labs) by 8.2% to 14.8% (Table 14).

**Table 14. Comparison of Proportions of Individuals Identified with T2DM using the eMERGE Primary Algorithm (does not include laboratory test results) versus the eMERGE Secondary Algorithm (includes laboratory test results)**

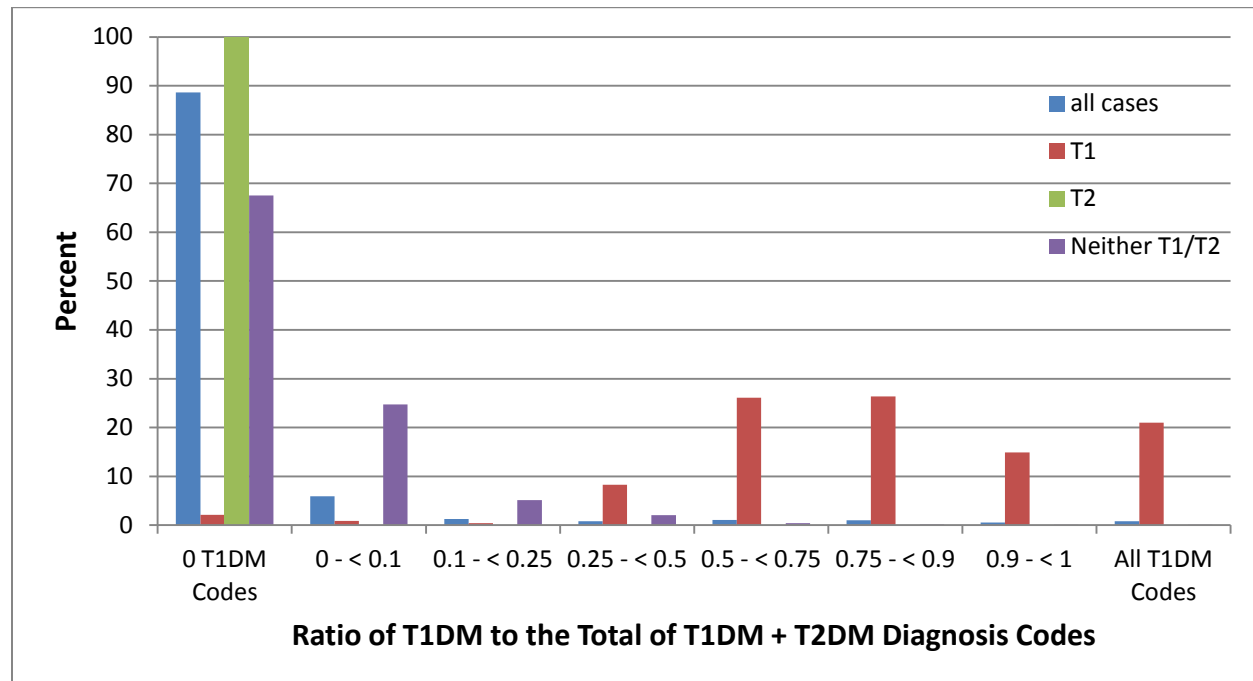
Site	Laboratory Results Data in MSDD?	eMERGE Primary Algorithm (no Labs) (%)	eMERGE Secondary Algorithm (with Labs)(%)	Additional Proportion of Patients Identified with T2DM Using Algorithm with Labs (%)
1	Yes	53.3	61.5	8.2
2	No	49.6	49.6	0
3	Yes	59.0	73.8	14.8
4	Yes	61.5	76.1	14.6
5	Yes	59.6	73.3	13.7

This demonstrates that **applying T2DM algorithms at sites without laboratory results values in the MSDD will miss an important minority of T2DM cases at those sites because a proportion of T2DM cases do not have the coded diagnosis and/or medication dispensings for T2DM required by the eMERGE Algorithms.**

## 2. Why Individuals Did Not Meet the Klompas Optimized T1DM Algorithm Diagnosis Ratio Criteria

Two of the Klompas Optimized T1DM Algorithm Pathways (1 and 2) requires that the ratio of T1DM diagnosis codes to the sum of T1DM plus T2DM diagnosis codes exceed 0.5 as a criterion to be classified as having T1DM. To inform whether this minimum ratio requirement contributed to many individuals not being identified as having T1DM, we examined the distribution of ratios of T1DM to T1DM plus T2DM codes among all 737,122 *Any DM* cases in the MSDD. As shown in Figure 4, more than 97% of individuals who did not qualify as having either T1DM or T2DM using the T1DM and T2DM algorithms had no T1DM (67.5%) code, a code ratio < 0.1 (24.7%), or a code ratio between 0.1 and < 0.25 (5.1%). Therefore, with a few exceptions (see Section VI.E. 3. below related to including the laboratory test criteria), the requirement for a ratio of T1DM to sum of T1DM plus T2DM diagnosis codes to exceed 0.5 did not substantially contribute to individuals not being identified with T1DM.

**Figure 4. Ratios of T1DM to T1DM plus T2DM Diagnosis Codes across All Individuals from Five Sites in the Mini-Sentinel Distributed Database with Any DM<sup>a</sup>**



<sup>a</sup> T2DM cases identified using the eMERGE Secondary Algorithm by definition had no T1DM codes

### 3. Exploration of Including C-Peptide and Diabetes Autoantibody Laboratory Test Results as Criteria for T1DM (Originally Pathways of Klompas Optimized Algorithm)

As noted previously, neither the MSDD nor the SUPREME-DM DataLink has C-peptide or diabetes autoantibody laboratory test result values. T1DM is defined by the presence of one or more of the diabetes autoantibody markers, including autoantibodies to glutamate decarboxylase 65 (GAD65), autoantibodies to insulin, islet cell autoantibodies, autoantibodies to the tyrosine phosphatases IA-2 and IA-2b, and autoantibodies to zinc transporter 8 (ZnT8).<sup>42</sup> C-peptide is produced by the  $\beta$ -cells in the pancreas when proinsulin splits apart and forms one molecule of C-peptide and one molecule of insulin. Because C-peptide and insulin are produced at the same rate, C-peptide is used as a marker of endogenous insulin production; low or undetectable C-peptide levels are consistent with T1DM. To help us better understand the potential consequences of not having these laboratory test results available, we set out to determine whether these laboratory test results are obtained in usual ambulatory care settings in the United States and, if obtained, whether the results identified additional T1DM cases in the MSDD. We examined all C-peptide and diabetes autoantibodies (including GAD65 antibody, insulin antibody, pancreatic islet cell antibody, and pancreatic islet cell IGG) laboratory test results for adults with Any DM in the MSDD at KPCO (N = 57,189), a representative ambulatory care setting that is the lead site for this project.

Among the adults with Any DM in the MSDD at KPCO, we found 2,798 C-peptide or diabetes autoantibody test results in the laboratory source data. These tests were completed within the study timeframe (January 1, 2006 - June 30, 2014). That is, these tests were completed after the patients reached adulthood and were most likely obtained to assess possible adult-onset T1DM, not possible juvenile-onset T1DM. C-peptide was the most common test with 1,821 results. GAD65 was next in

frequency with 729 results, and taken together insulin antibody, pancreatic islet cell antibody, and pancreatic islet cell IGG accounted for 248 additional test results. More than one test was often conducted for the same individual. The 2,798 results were from 1,709 unique individuals. This confirmed that C-peptide or diabetes autoantibody laboratory tests for T1DM in adults are obtained in usual care in the United States.

The Klompas Optimized T1DM Algorithm had identified 2,724 individuals at KPCO as having T1DM; 608 of these individuals (22.3%) also had C-peptide negative and/or diabetes autoantibody positive laboratory test results. Importantly, the results of these T1DM laboratory tests identified an *additional* 226 adults in the MSDD at KPCO as having T1DM, increasing the proportion of patients identified with T1DM at KPCO from 4.8% (2,724 of 57,189) to 5.2% (2,950 of 57,189). The numbers of specific laboratory test results indicating T1DM and the number and percentages of those individuals identified using the Klompas Optimized T1DM Algorithm are shown in Table 15.

**Table 15. C-peptide Negative and Diabetes Antibody Positive Laboratory Test Results at One Site among Patients Identified as Having Any DM in the MSDD**

T1DM Laboratory Test Type	Number of Individuals with Laboratory Results Consistent with T1DM (C-peptide negative; diabetes autoantibody positive) <sup>a</sup>	Number (%) of Individuals with Laboratory Results Consistent with T1DM Identified using Klompas Optimized T1DM Algorithm <sup>a</sup>	Number (%) of Individuals Not Identified using Klompas Optimized T1DM Algorithm <sup>a</sup>
C-peptide	648	466 (71.9)	182 (28.1)
GAD 65 antibody	215	136 (63.3)	79 (36.7)
Insulin antibody	24	21 (87.5)	3 (12.5)
Pancreatic islet cell antibody	1	1 (100)	0 (0)
Pancreatic islet cell IGG	37	27 (73.0)	10 (27.0)
Total unique individuals	752	526 (69.9)	226 (30.1)
<sup>a</sup> Several laboratory test types could have been conducted on a single individual, therefore the sum across laboratory test types does not equal the number of unique individuals			

In examining the electronic data for the 226 individuals newly-identified with T1DM using the laboratory test results, 182 (80.5%) had not met criteria for either T1DM or T2DM using the Klompas T1DM Optimized or eMERGE T2DM Secondary Algorithms (i.e., they had DM of uncertain type) and 44 (19.5%) had met criteria for T2DM using the eMERGE T2DM Secondary Algorithm. The only medication used by 97 of these individuals was insulin, but these 97 did not have a ratio of T1DM to T1DM plus T2DM diagnosis codes > 0.5 as required for meeting criteria using the T1DM Klompas Optimized Algorithm. Similarly, 21 additional individuals had a dispensing of glucagon, but they did not have a majority of T1DM diagnosis codes. Thirty-one individuals had a majority of T1DM diagnosis codes, but had been dispensed an oral antidiabetic medication other than metformin. ***This exploration confirms the value of C-peptide and diabetes autoantibody laboratory test results in augmenting the case identification of T1DM.***

## VIII. RECOMMENDATIONS FOR USING THE SUPREME-DM DATALINK FOR ANY DM, T1DM, AND T2DM

### Recommendation 1: Employ the SUPREME-DM DataLink as the Registry of Adults with Any DM for Sentinel Public Health Surveillance Activities

The workgroup recommends that the SUPREME-DM DataLink be used as the gold standard for adults with Any DM. Using patients in the SUPREME-DM DataLink can replace medical record review as the alternative reference source for cases with Any DM. In addition to efficiencies gained from the fact that individuals in the SUPREME-DM DataLink are already linked to individuals in the MSDD, there are other advantages to employing the SUPREME-DM DataLink as the gold standard DM registry in medical product surveillance. A few of these include that the SUPREME-DM DataLink investigators and programming teams are productive and collaborative, many data elements available in the SUPREME-DM DataLink are not available in the MSDD such as multiple types of laboratory test results, social behavioral data, race and ethnicity, cause of death, and many years of follow-up data are available for most adults in the SUPREME-DM DataLink.

To optimize the usefulness of the SUPREME-DM DataLink as the gold standard registry for DM in Sentinel, the workgroup also recommends the SUPREME-DM DataLink be updated at least yearly. The SUPREME-DM DataLink from the five participating sites identified 737,122 adults with Any DM between January 1, 2006 and June 30, 2014). Updating the SUPREME-DM DataLink at these five sites will identify an additional 50,000 – 60,000 adults with Any DM each year and lengthen follow-up most individuals already in the SUPREME-DM DataLink.

### Recommendation 2: Consider the SUPREME-DM DataLink as the Registry of Adults with T1DM

The adult T1DM cases identified by the Modified Klompas T1DM Algorithm in the SUPREME-DM DataLink are considered by the workgroup to be true cases. ***The workgroup believes the T1DM algorithm tested by the workgroup identified a higher number of adults in the DataLink as having T1DM than are available in any other T1DM registry.*** The concern was under-ascertainment of cases due to lack of access to C-peptide and diabetes autoantibody results. In this project we demonstrated that C-peptide and diabetes autoantibody laboratory tests for T1DM in adults were available at a participating SUPREME-DM site and confirmed that including these laboratory test results enhanced T1DM case identification (increasing the proportion of cases from 4.8% to 5.2% of the SUPREME-DM DataLink Any DM population at that site). Other SUPREME-DM sites that participated in this project are now adding C-peptide and diabetes autoantibodies laboratory test results to the SUPREME-DM DataLink registry. Therefore, the Modified Klompas T1DM Algorithm *that includes laboratory test result criteria* can be used to identify patients in the SUPREME-DM DataLink Registry with adult T1DM. ***Enhancing the SUPREME-DM DataLink with these T1DM laboratory test results removes the current shortcoming to considering the SUPREME-DM DataLink as the gold standard for T1DM*** (and these laboratory test results can be added more efficiently to the SUPREME-DM DataLink than to the MSDD).

### Recommendation 3: Consider the SUPREME-DM DataLink the Gold Standard Registry of Adults with non-T1DM

**Recommendation 4: Consider Testing the Newly-Developed T2DM Algorithm in the SUPREME-DM DataLink Any DM Population Prior to Considering the SUPREME-DM DataLink the Gold Standard Registry of Adults with T2DM**

Because the T2DM algorithms tested in this project only identified 70.7% of patients with Any DM as having T2DM, we do not recommend either tested algorithm (see Section IX.C.). We also do not recommend applying either of those tested algorithms to the SUPREME-DM DataLink population to identify the T2DM population. Instead, we recommend the approaches detailed below.

*The workgroup recommends employing the SUPREME-DM DataLink as the gold standard non-T1DM registry for adults*, that is, including adults with T2DM and the relatively small number of patients that have DM of uncertain or rare types. Nearly all adults that do not have T1DM have T2DM. For most Sentinel DM-related safety surveillance activities, the medical products of interest are used similarly in patients with T2DM and DM of uncertain or rare types and differentiating between T2DM and DM of uncertain or rare types is usually unnecessary. By first identifying individuals in the SUPREME-DM DataLink with T1DM using the Modified Klompas T1DM Algorithm, those individuals can be removed from consideration and all remaining individuals can be classified as having *non-T1DM*.

The workgroup developed a new T2DM algorithm that retains most diagnosis and medication criteria from the tested algorithms and added criteria to identify additional patients with T2DM (See Section IX.C.1. for details of the Pathways in this new algorithm). ***To increase confidence that the SUPREME-DM DataLink can be used as a gold standard T2DM registry, we recommend testing the newly-developed algorithm in the SUPREME-DM DataLink Any DM population, with validation through review of a subset of medical records. This work could efficiently yield two work products: a validated T2DM algorithm (where none currently exists) and the SUPREME-DM DataLink as the gold standard T2DM registry.***

## IX. RECOMMENDED ALGORITHMS FOR IDENTIFYING ANY DM, T1DM AND T2DM IN THE MSDD POPULATION

### A. IDENTIFYING ANY DM IN THE MSDD POPULATION

The MSDD includes patients from *multiple* sites, including a large number of insurers, broad geographic regions, and many care delivery settings. As a consequence, applying a DM identification algorithm tested in a *single* site DM registry may not correctly classify MSDD patients. Strengths of this work include successfully linking patients from a *multi-site any* DM registry – the SUPREME-DM DataLink – to patients in the MSDD and then testing the SE and PPV of several published algorithms using the SUPREME-DM DataLink as the gold standard. All tested *Any* DM algorithms performed very well compared to the gold standard SUPREME-DM DataLink criteria, lending credibility to the three tested algorithms, as well as further confirming the SUPREME-DM DataLink Algorithm is the gold standard *Any* DM algorithm. ***We recommend the original SUPREME-DM DataLink Algorithm criteria to identify patients with Any DM in the MSDD which includes laboratory results when available from the Data Partner.***

**Recommendation 1: Consider the SUPREME-DM DataLink Algorithm the Gold Standard to Identify Adults with Any DM in the MSDD**

- $\geq 1$  inpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41,
- **OR**  $\geq 2$  of any of the following (when the two events are from the same source [e.g. two outpatient diagnoses or two elevated laboratory values], they must occur on separate dates  $\leq 730$  days apart)
  - Outpatient ICD-9 codes from among the following: 250.XX, 357.2, 362.01-362.07, 366.41
  - Antidiabetic medication (two dispensing of metformin or two dispensings of thiazolidinediones with no other indication of diabetes are not included)
  - A1c  $\geq 6.5\%$
  - Fasting plasma glucose  $\geq 126$  mg/dl
  - Random plasma glucose  $\geq 200$ mg/dl
- **Exclude** criteria ascertained during periods of pregnancy to ensure gestational diabetes is not inadvertently captured

While the recommended SUPREME-DM DataLink algorithm includes diagnosis, medication, and laboratory results criteria, we also confirmed the utility of applying the SUPREME-DM DataLink Algorithm modified without laboratory results (identified 5% fewer individuals as having Any DM in the MSDD across the five participating Sentinel Data Partner sites). It was important to confirm the performance of the gold standard Any DM algorithm modified to include only diagnosis and medication criteria because some Sentinel Data Partners either have no or only partial access to laboratory test results for their enrollees. Some under-ascertainment of Any DM at Data Partner sites with incomplete or no laboratory test results will invariably occur, but this is unavoidable regardless of the criteria/algorithm used when laboratory results are not available. Thus, **we recommend employing the gold standard Any DM algorithm at all Data Partner sites, whether claims-based or integrated delivery systems.** In other words, the critical issue is not whether the Data Partner is claims-based or an integrated delivery system, but rather how complete the data are on the individual. **As long as the healthcare system has pharmacy claims and inpatient and outpatient diagnosis claims, the algorithm is expected to perform well** (e.g., the results are generalizable).

**B. CRITERIA/ALGORITHM(S) FOR IDENTIFYING T1DM IN THE MSDD POPULATION**

Most published T1DM algorithms were designed to identify T1DM in children and youth.<sup>36,39,45</sup> The only two T1DM algorithms we found that had been tested in adults were both by Klompas.<sup>4</sup> ***Prior to applying a Klompas algorithm to identify patients with T1DM in the MSDD population, we recommend that individuals first meet the SUPREME-DM DataLink Algorithm gold standard Any DM criteria because the Klompas T1DM algorithms were designed to determine patients with T1DM within a population already identified as having Any DM.***<sup>4</sup>

After testing both Klompas algorithms, we determined the Klompas Optimized T1DM Algorithm identified more cases as having T1DM in part because of a pathway (Pathway 2) that identified patients with diagnosis + glucagon dispensing and in part because it allowed patients who had ever been



prescribed a medication other than insulin or metformin to be considered T1DM (such as adults with new-onset T1DM who were started on multiple oral agents before they fail.) **We recommend a Modified Klompas T1DM Algorithm to identify patients with T1DM in the MSDD.** This Algorithm is expected to perform similarly in both Data Partners that are claims-based and those that are integrated delivery systems.

The Klompas Optimized T1DM Algorithm originally included C-peptide negative (Pathway 4) or DM autoantibodies positive (Pathway 5) pathways. Neither of these laboratory test results types is available in the MSDD. We therefore had to modify the Klompas Optimized Algorithms prior to testing it in the MSDD Any DM population. Because definitive existence of T1DM in adult onset T1DM is made by the presence of diabetes antibodies or a low or negative C-peptide value, at the lead site we examined whether these laboratory tests were used and if so, whether or not their use increased the proportion of patients identified with T1DM. We confirmed that these laboratory tests were obtained in many patients who had T1DM based on the Klompas Optimized Algorithm. Further, we determined that additional patients would have been identified with T1DM had these laboratory results been available in the MSDD: the proportion of patients identified with T1DM would have increased from 4.8% to 5.2% of the Any DM population at that data partner site. **In the absence of C-peptide and DM autoantibody laboratory test results criteria, the recommended Modified Klompas T1DM Algorithm will underascertain T1DM cases at all data partner sites.**

#### **Recommendation 2: Employ the Modified Klompas T1DM Algorithm to Identify Adults with T1DM in the MSDD**

- Before applying this algorithm, individuals must meet the criteria for the Any DM algorithm (see Recommendation 1)
- Pathway 1:
  - Diagnosis codes: a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes >0.5, **AND**
  - Medications: insulin or metformin is allowed, but any other antidiabetes medication results in exclusion
- **OR** Pathway 2:
  - Diagnosis codes: a ratio of type I (ICD-9 250.X1 or 250.X3) to type II (ICD-9 250.X0 or 250.X2) codes >0.5, **AND**
  - Medication: glucagon

The Klompas Optimized T1DM Algorithm originally included a Pathway 3 that identified individuals as having T1DM based solely on a single dispensing of urine acetone test strips. **We do not recommend including a single dispensing of urine acetone test strips as a criterion in the Modified Klompas T1DM algorithm because it lacks face validity.** Medical record review of cases that met criteria for the Klompas Optimized T1DM Algorithm based on a dispensing of urine acetone test strips should be done prior to further considering this criterion.



**Recommendation 3: More Complete Identification of T1DM in Adults in the MSDD**

Applying the Modified Klompas T1DM Algorithm without laboratory test results under-ascertains T1DM cases. If more complete identification of T1DM cases in adults is an FDA priority, we recommend that C-peptide and diabetes autoantibody laboratory results be added to the MSDD and that either a negative C-peptide or positive diabetes autoantibodies be considered as diagnostic of T1DM. These laboratory test results could be added from both claims-based Data Partners that have laboratory results available as well as integrated delivery systems with laboratory results, because their use would be to *enhance* identification of adults with T1DM, and thereby reduce misclassification.

One caution with recommending the Modified Klompas T1DM Algorithm (Pathways 1 and 2 without laboratory results data Pathways 4 and 5) is that the proportion of adults identified with T1DM varied across the five participating sites from 2.4% to 7.4% of the *Any DM* populations. Examination of the median ages of patients with *Any DM* from those sites did not help explain the variation. While there may be systematic differences in how the Modified Klompas T1DM Algorithm performed at the sites with the lowest and highest proportions of patients identified with T1DM, we cannot explain this variation.

**C. CRITERIA/ALGORITHM(S) FOR IDENTIFYING T2DM IN THE MSDD POPULATION**

***We do not recommend either of the tested eMERGE T2DM algorithms.*** The eMERGE algorithms are unsuitable when laboratory results data are not available, they are cumbersome to use, and the better of the two eMERGE algorithms still only identified 70.7% of individuals with *any DM* as having T2DM. Therefore, we recommend other options (see below) for identifying patients as having T2DM. Our recommendations have varying strengths and weaknesses in terms of ease of use and in level of confidence that the individuals identified actually have T2DM (e.g., likely misclassification). Prior to applying any algorithm to identify patients with T2DM in the MSDD population, we recommend that individuals meet the gold standard *Any DM* criteria previously discussed.

**Recommendation 4: Tested T2DM Algorithms are Not Recommended. Recommended Options Include:**

- T2DM Algorithm Option 1
  - A newly-developed algorithm that retains most diagnosis and medication criteria from the tested algorithms and adds criteria to identify additional patients with T2DM.
  - Minimizes case misclassification, but is somewhat complex to implement because of multiple criteria sets.
- *Non-T1DM* “Algorithm” Option 2
  - Identify and exclude individuals with T1DM using the recommended T1DM algorithm. Classify all remaining individuals as T2DM.
  - Easy to implement, but misclassifies a percentage of cases because individuals with DM of uncertain type are considered T2DM.

## 1. Recommended Option 1: Identifying T2DM in the MSDD Population

Based on further exploration of the individuals who did not meet the tested T2DM algorithm criteria, we recommend a set of criteria revised and modified from the two tested eMERGE algorithms. These revised and modified criteria are expected to identify 83% to 90% of adult patients with *Any* DM as having T2DM. Because the algorithm we recommend for identifying T2DM has been modified from the tested algorithms, it is newly-developed, and as yet, unnamed. At the discretion of the FDA, the algorithm could be entitled the “Sentinel-SUPREME-DM DataLink T2DM Algorithm.” The criteria for this new T2DM Algorithm retain most diagnosis and medication criteria from the eMERGE algorithms, add criteria for patients with T2DM who are not treated with medication, who do not have laboratory test results in the MSDD, and whose initial treatment included a combination of oral antidiabetic medications. The criteria for this new T2DM Algorithm also clarify acceptable care settings from which T2DM ICD-9 codes can be used, modify the blood glucose and HbA1c criteria, and require a maximum timeframe within which the criteria must all be met.

We recommend the following sets of criteria (Pathways 1 - 5) as the algorithm to identify patients with T2DM in the MSDD. Taken together, all five Pathways and criteria sets constitute a single algorithm to identify patients with T2DM in the MSDD (i.e., patients can enter through any of Pathways 1, 2, 3, 4, or 5):

- Pathways 1 – 5: No T1DM ICD-9 codes (250.x1, 250.x3) from any setting
- Pathways 1-4: Require T2DM ICD-9 codes (ICD-9 codes 250.x0, 250.x2; excluding 250.10, 250.12) from inpatient, outpatient, or ED visits
- Pathways 1 - 5: All criteria must be met within 730 days of each other
  - Pathway 1
    - > 2 T2DM ICD-9 codes, AND
    - No insulin or T2DM medication
  - OR Pathway 2
    - > 1 T2DM ICD-9 codes, AND
    - Any T2DM medication, AND
    - No insulin
  - **OR Pathway 3**
    - $\geq 1$  T2DM ICD-9 codes, **AND**
    - Any T2DM medication, **AND**
    - Any insulin, **AND**
    - Date of T2DM medication < date of insulin
  - **OR Pathway 4**
    - $\geq 1$  T2DM ICD-9 codes, **AND**
    - No insulin, **AND**
    - No T2DM medication, **AND**
    - Random plasma glucose  $\geq 200$  mg/dL, **OR** fasting plasma glucose  $\geq 126$  mg/dL, **OR** HgA1c  $\geq 6.5\%$
  - **OR Pathway 5**
    - Any T2DM medication, AND
    - Random plasma glucose > 200 mg/dL, OR fasting plasma glucose > 126 mg/dL OR HgA1c > 6.5%

The strengths of this recommended algorithm are that it is based on criteria known to identify a cohort of patients who do have T2DM, does not include many patients with diabetes of uncertain or rare types (e.g., LADA, MODY), and does not inadvertently include patients with T1DM. The limitation of this algorithm is that it requires somewhat complex programming code to operationalize and it is somewhat unwieldy to explain. Importantly, this algorithm should be used when it is important to have relatively high confidence that the cohort of patients identified to study a drug safety question includes only patients with T2DM (i.e., anticipated to have high specificity).

## 2. Recommended Option 2: Identifying *Non-T1DM* in the MSDD Population

A simpler approach is to identify *Non-T1DM* in the MSDD population. This is accomplished by first identifying individuals with *Any DM*, then identifying and excluding individuals with T1DM using the Modified Klompas T1DM Algorithm, and finally, classifying all remaining individuals as having T2DM (e.g., just excluding individuals with T1DM). The strength of this approach is ease of implementation. The limitation of this approach is that the cohort includes some patients who do not have T2DM; it allows patients who have DM of uncertain type and T1DM patients from sites without laboratory results data to remain in the *Non-T1DM* cohort. An example of an appropriate use of this approach would be when the drug safety question involves a drug used in patients with DM that is not T1DM and there is little concern the outcome of interest differs by T2DM, DM of uncertain type, or DM of other rare types.

## D. IMPLICATIONS FOR SENTINEL ROUTINE TOOLS

### *Implications for Sentinel Routine Query Tools*

This workgroup focused on utilizing the SUPREME-DM DataLink (or registry) to explore algorithms to identify cohorts of individuals with ‘*Any DM*’, ‘*T1DM*’, and ‘*T2DM*’ within the Sentinel Distributed Database. Although Workgroup aims were accomplished with *de novo* code, it would be possible to utilize Sentinel routine tools for algorithm implementation if recommended parameters are modified to include index dates. Sentinel routine tools were designed in the context of medical product safety surveillance, and require use of index dates to identify cohorts and health outcomes of interest. They utilize inclusion/exclusion criteria for cohort selection, which are assessed during a requester defined number of days before, on, or after the exposure episode index date. Similarly, Sentinel tools also require use of index dates to identify specific health outcomes of interest. Although the workgroup did not focus on identifying incident DM outcomes subsequent to a medical product exposure, it would be possible to modify recommended algorithm parameters (e.g., specify index dates) to capture *Any DM* and T2DM outcomes.

- Algorithms for *Any DM* and T2DM may be able to be implemented in the current Sentinel tools if all parameters and temporal relationships are defined in a tool-specific manner. For example, as algorithms for *Any DM* exclude pregnancy, an already developed algorithm for pregnancy would need to be adapted to the Sentinel tool framework. NDC lists for relevant antidiabetic agents have already been developed by the Workgroup and, similar to use of other NDC lists, would need periodic updates. All examined DM algorithms include several criteria (or criteria sets), and temporal relationships are at times, integral. Thus, timeframes included within each algorithm would need to be reinterpreted in terms of index dates. Current Sentinel tools are able to utilize the glucose and glycosylated hemoglobin (HbA1c) laboratory result values included in the Sentinel Common Data Model; it is possible to distinguish elevated versus non-elevated random and fasting glucose values and HbA1c at Data Partners that contribute laboratory result values

to the laboratory results table. Thus, the individual criteria for *Any DM* and T2DM (if reinterpreted) appear to be compatible with Sentinel tools.

- To implement algorithms for T1DM examined in this report, Sentinel tools would need to be modified to accommodate ratios of T1DM to T2DM codes. However, should accommodation of these T1DM algorithms be deemed a priority, it would be possible to determine if the tools can be updated to accommodate these algorithms.

In summary, algorithms for *Any DM* and T2DM described in this report appear to be compatible with current Sentinel tools, but algorithms for T1DM currently cannot be implemented using existing Sentinel tools. The Sentinel Operations Center recommends piloting algorithm implementation to confirm the *Any DM* and T2DM algorithm parameters can be implemented as intended using the existing query tools. This will assist with identifying any gaps between current and needed tool functionality. Careful consideration of potential medical product safety question (s) of interest during this pilot will also help to ensure algorithms are implemented appropriately.

## X. RECOMMENDATIONS FOR FUTURE WORK

### A. FURTHER ENHANCEMENT OF THE SUPREME-DM DATALINK

#### 1. Ongoing Refresh of the SUPREME-DM DataLink Database

Given that the workgroup successfully linked essentially 100% of the adult patients in the SUPREME-DM DataLink to adult patients in the MSDD from January 1, 2006 through June 30, 2014 (Specific Aim 1), and confirmed that the SUPREME-DM DataLink performed excellently as the gold standard registry for adults with *Any DM* (Specific Aim 2), we have recommended use of the SUPREME-DM DataLink resource as the gold standard registry in future DM projects. Some of the advantages to employing the SUPREME-DM DataLink as the gold standard include that the it is comprised of over 737,000 adults with *Any DM* already linked to the MSDD, contains a large cohort of adults already confirmed with T1DM, enables access to additional data elements for these patients beyond those available in the MSDD (e.g., additional laboratory test result types, social behavioral data, race and ethnicity, cause of death), and 3) includes years of healthcare follow-up data for most SUPREME-DM Datalink patients. ***To further enhance longitudinal follow-up of patients already in the SUPREME-DM DataLink and to increase the number of adults with Any DM in the gold standard population in an ongoing manner, the workgroup recommends the SUPREME-DM DataLink be refreshed yearly.***

#### 2. Test the Newly-developed T2DM Algorithm in the SUPREME-DM DataLink *Any DM* Population

See Section VIII for details of this recommended enhancement. See Sections IX. and IX.C. for information about this new algorithm.

### B. FURTHER IMPROVEMENTS TO THE SE AND PPV OF THE ALGORITHMS RECOMMENDED FOR USE IN THE MSDD POPULATION

Although we have completed this work to develop or refine and automate algorithmic approaches to replace medical record review for identifying patients with *Any DM*, T1DM and T2DM, as anticipated,

this work was incremental towards validating DM HOI of interest, and some questions remain. For example, because no gold standard T2DM algorithm existed, the workgroup was only able to compare case identification between two T2DM algorithms, without knowing “truth.” As an Addendum, (similar to the originally recommended Specific Aim 3) we recommend reviewing a sample of medical records from selected subgroups of patients. ***By reviewing medical records for patient subgroups where uncertainty remains, we can avoid the need to justify foregoing medical record review in future projects that seek to identify patients in the MSDD (or the SUPREME-DM DataLink) with Any DM, T1DM, or T2DM. This review of selected medical records now will increase confidence in the SE and PPV of the automated algorithms we recommend and lend credibility to future automated DM-related Active Risk Identification and Assessment (ARIA) activities. The workgroup’s medical record review recommendations can be adapted to FDA priorities.***

1. **If the priority is to ensure DM algorithms applied in the MSDD are refined to ensure high SE and/or PPV of T1DM or T2DM, the workgroup recommends:**
  - **Chart review of patients with DM of uncertain type. This group includes, for example, patients with a mixture of T1DM and T2DM diagnosis codes.**
  - Chart review will aid understanding of whether these patients have DM of uncertain type or if further algorithm refinement would classify them as T1DM or T2DM. The public health benefit is likely greatest in two situations: a) additional patients that are correctly classified as T1DM given that the currently-recommended algorithm under-ascertains T1DM cases, and b) some patients will be identified as having conditions such as LADA or MODY (i.e., better understandings of how such patients are coded).
  
2. **If identifying a very “clean” cohort of individuals with T1DM in the MSDD (i.e., anticipated to have high specificity) is the key priority, the workgroup recommends:**
  - **Reviewing charts for patients identified using the Optimized Klompas T1DM Algorithm as modified for testing in the MSDD by the workgroup (Pathways 1 - 3). Specifically, the workgroup recommends chart review for patients who met algorithm criteria by having a dispensing of glucagon and did not have insulin dispensing but who may have been dispensed antidiabetic medications typically used for T2DM, or who met algorithm criteria by only having a dispensing of urine acetone test strips.**
  - Confirming whether or not this subgroup of patients do have T1DM will inform whether further algorithm refinement is needed, thus optimizing FDA investment in the T1DM algorithm work.
  
3. **If identifying patients with T2DM in the MSDD is a key priority, the workgroup recommends:**
  - **Validating the newly-developed MSDD-specific T2DM algorithm. The workgroup validated the eMERGE Algorithms as part of this project and found they did not perform well. The workgroup extensively modified those existing algorithms and recommended a newly-developed algorithm to identify T2DM cases.**
  - The new algorithm the workgroup recommends has not been validated. While the workgroup developed the new algorithm based in part on best available evidence from tested algorithms, validating the new algorithm is crucial before broadly applying it to public health surveillance activities. Depending on priority, part of validating this new T2DM validation could include medical chart review for any of the following subgroups: a) Individuals identified with the new T2DM algorithm who would not be identified with the

eMERGE Algorithms, b) individuals who could not be classified as either T1DM or T2DM, and c) individuals meeting algorithm criteria due to ED visits.

4. **If identifying additional patients with Any DM in the MSDD is a priority, the workgroup recommends:**

- **Examining charts for patients who would meet criteria for Any DM as a result of expanding the diagnosis criteria to include ED visits.**
- Determining whether ED visits are appropriate to include as a setting from which diagnosis codes should be ascertained will clarify whether the recommended algorithms (e.g., the new T2DM algorithm and the Any DM algorithm) should include ED visits as part of case definitions.

5. **Rationale for the above recommendations**

- DM is a high priority health outcome of interest to the FDA in postmarketing active medical product safety surveillance and in protocol-based assessments. Recent examples include:
  - A recommendation that newly approved medications for treatment of diabetes be thoroughly and systematically evaluated for cardiovascular risk.<sup>47</sup>
  - T2DM was the cohort of interest for the Mini-Sentinel acute myocardial infarction and antidiabetic agents protocol.<sup>48</sup>
  - T2DM was the outcome of interest for the Metabolic Effects of Second Generation Antipsychotics in Youth, Subprojects 1, 2, and 3.<sup>49</sup>
  - The Protocol Core (now called the Applied Surveillance Core) was charged with conducting two literature reviews on DM, including a review of T1DM as an outcome and a review of persons with T1DM or T2DM.<sup>2,50</sup>
  - The Mini-Sentinel Alternative Methods for health outcomes of interest highlighted DM.<sup>7</sup>
  - The Rivaroxaban and Immunoglobulin Administration and Thromboembolic Events protocols use DM as a covariate.<sup>51,52</sup>
  - Mini-Sentinel modular programs have focused on DM and laboratory test results.<sup>53</sup>
  - A White Paper addressing methods to evaluate the impact of FDA regulator actions highlighted a diabetes drug.<sup>54</sup>
- Medical record reviews of a subset of individuals in the selected priority subgroups will provide additional information to FDA to guide future medical product safety work and ARIA activities. The existence of patients with DM in these subgroups will be clarified, the presence of such patients in the MSDD will be better addressed (e.g., characteristics associated with discrepancies and uncertainties in the MSDD will be known), and as necessary, DM automated algorithms can be further refined.
- It is important to avoid further misclassification of patients in T1DM and T2DM algorithms, as misclassification contributes to bias in analyses utilizing these algorithms. For electronic claims data to be used for medical product safety work and ARIA activities, a very high proportion of people identified with T1DM or T2DM should actually have that DM type. Therefore, we seek to further avoid misidentification of patients in T1DM and T2DM algorithms.
- Conducting medical record reviews within the context of the SUPREME-DM project will be efficient. Participating workgroup sites have extensive experience conducting medical record review using trained personnel employed by the sites' research departments. All sites have on-site access to EHR that include not only comprehensive ambulatory care



information, but also information such as discharge summaries from care rendered in hospital and ED environments.

**6. Possible deliverables upon completion of the review of the sample of medical records**

- An addendum report that clarifies whether the algorithms recommended in the main project report should be used as recommended or whether further modification will improve the accuracy of handling these patient types for future projects.
- A manuscript suitable for submission to a peer-review journal for publication based on the T1DM algorithm work. Completion of this manuscript would require completing incorporating C-peptide and diabetes autoantibody lab values as was done at KPCO (currently underway at some SUPREME-DM Sites), chart review of up to 200 charts of T1DM cases and possibly chart review of up to 200 charts of uncertain DM individuals based on the T1DM and T2DM algorithms.
- The chart abstraction tool(s) used in the medical record review (if of interest to the FDA).

**C. ADDING C-PEPTIDE AND DIABETES AUTOANTIBODIES LABORATORY TEST RESULTS TO THE MSDD TO IMPROVE CASE IDENTIFICATION OF ADULT-ONSET T1DM**

If more complete identification of T1DM cases in adults in the MSDD is a priority for the FDA for active medical product safety surveillance, we recommend C-peptide and diabetes autoantibody laboratory test results be added to the MSDD (and Pathways 4 and 5 added to the Modified Klompas T1DM Algorithm). Reasons for these suggested additions to the MSDD Laboratory Results Table are listed below (these reasons also apply to the ongoing work at the SUPREME-DM sites that are adding these laboratory tests to the DataLink).

1. **T1DM is caused by cellular-mediated autoimmune destruction of the pancreatic  $\beta$ -cells.** This typically leads to little or no pancreatic insulin secretion, which can be detected as a low or undetectable C-peptide level. Individuals with positive diabetes autoantibodies, by definition have T1DM. T2DM is primarily a disease of relative (as opposed to absolute) insulin deficiency. The vast majority of individuals with undetectable C-peptide levels have T1DM.
2. **Pilot work with C-peptide and diabetes autoantibody laboratory test results confirmed that T1DM laboratory tests are obtained in adults in usual care settings in the United States.** Although testing occurred in a small proportion of patients, it was clear the patients tested were those clinicians suspected of having T1DM, as evidenced by the fact that 56% of the patients in whom these laboratory tests were obtained had one or more test results consistent with T1DM.
3. **Incorporating C-peptide and diabetes autoantibody results into the MSDD Laboratory Results Table for patients identified with *Any* DM would enhance case identification of T1DM across all sites with laboratory results data available.**
  - C-peptide and diabetes autoantibody tests are often conducted in the outpatient care environment and thus, even data partner sites with only outpatient laboratory results can enhance their T1DM case identification through inclusion of these laboratory test results when the T1DM algorithm is applied.
  - We are likely missing T1DM laboratory test results for individuals whose T1DM was identified during childhood, but for those individuals, a T1DM diagnosis code is likely to have been consistently applied. Therefore, the laboratory tests are most likely to be useful in identifying adults with new-onset T1DM.



4. **Even using a T1DM algorithm (the T1DM algorithm without laboratory test results) it is not always possible to separate T1DM from T2DM. This results in many new-onset adult T1DM cases not being accurately characterized.**
  - Accurate separation of *Any* DM into T1DM and T2DM is important for “routine” postmarketing active medical product safety surveillance (see above examples of DM as a high priority health outcome of interest to the FDA in postmarketing active medical product safety surveillance and in protocol-based assessments)

#### **D. ALGORITHM DIAGNOSIS CODE CONVERSION TO ICD-10**

The diagnosis codes in the algorithms recommended by the workgroup are ICD-9 codes. With the United States healthcare system transitioning to ICD-10 coding, for the recommended algorithms to remain usable in the future (e.g., in ARIA activities), the diagnosis codes in the current algorithms need to be mapped from ICD-9 to ICD-10. Programs are available that map individual codes from ICD-9 to ICD-10 (e.g., ICD-9 250.x codes). Clinical judgment is still required as matches are not always clear. If conversion to ICD-10 is a priority for the FDA, members of the workgroup have experience with using an ICD-9 to ICD-10 translation program. Also, clinical expertise is available within the workgroup to assist with mapping when questions arise.

## XI. TABLES AND FIGURES

### A. TABLES

Table 1. Probabilistic Match and Direct Match Gold Standard in Mini-Sentinel Distributed Database and SUPREME-DM DataLink Linkage Project .....	10 -
Table 2. <i>Any</i> Diabetes Algorithms .....	12 -
Table 3. Case Ascertainment of the Solberg Primary <i>Any</i> DM Algorithm versus the SUPREME-DM DataLink Algorithm Gold Standard .....	16 -
Table 4. Case Ascertainment with the Solberg Secondary <i>Any</i> DM Algorithm Modified to Include Emergency Department Visits versus the Gold Standard .....	18 -
Table 5. Case Ascertainment with the Zgibor Algorithm versus the Gold Standard .....	20 -
Table 6. Algorithms for Type 1 Diabetes Mellitus .....	23 -
Table 7. Patients Identified with T1DM in the MSDD Based on the Klompas Primary T1DM Algorithm .....	24 -
Table 8. Patients Identified with T1DM in the MSDD Based on the Klompas Optimized Algorithm...-	25 -
Table 9. Comparisons of the Klompas Primary T1DM Algorithm and the Klompas Optimized T1DM Algorithm with Neither Considered a Gold Standard .....	26 -
Table 10. Algorithms for Type 2 Diabetes Mellitus .....	28 -
Table 11. Patients Identified with T2DM in the MSDD Based on the eMERGE Primary Algorithm....-	29 -
Table 12. Patients Identified with T2DM in the MSDD Based on the eMERGE Secondary Algorithm-	30 -
Table 13. Comparisons of T2DM eMERGE Primary Algorithm (without Laboratory Test Results) and eMERGE Secondary Algorithm (with Laboratory Test Results) with Neither Considered a Gold Standard.....	31 -
Table 14. Comparison of Proportions of Individuals Identified with T2DM using the eMERGE Primary Algorithm (does not include laboratory test results) versus the eMERGE Secondary Algorithm (includes laboratory test results) .....	34 -
Table 15. C-peptide Negative and Diabetes Antibody Positive Laboratory Test Results at One Site among Patients Identified as Having <i>Any</i> DM in the MSDD.....	36 -

### B. FIGURES

Figure 1. Linked HOI alternative validation .....	9 -
Figure 2. SUPREME-DM DataLink Linkage to Mini-Sentinel Distributed Database Project Cohort for SUPREME-DM DataLink Algorithm Gold Standard for <i>Any</i> Diabetes.....	14 -
Figure 4. Ratios of T1DM to T1DM plus T2DM Diagnosis Codes across All Individuals from Five Sites in the Mini-Sentinel Distributed Database with <i>Any</i> DM <sup>a</sup> .....	35 -
Figure 5. Algorithm for identifying T2DM cases in the EMR .....	61 -
Figure 6. Algorithm for identifying T2DM controls in the EMR .....	61 -

## XII. REFERENCES

1. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroschikoski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. *Am J Med Qual* 2006;21:238-45.
2. Leonard CE, Freeman CP, Razzaghi H, et al. Mini-Sentinel methods: 15 cohorts of interest for surveillance preparedness. [http://www.mini-sentinel.org/work\\_products/Statistical\\_Methods/Mini-Sentinel\\_Methods\\_15-Cohorts-of-Interest-for-Surveillance-Preparedness.pdf](http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_15-Cohorts-of-Interest-for-Surveillance-Preparedness.pdf). Accessed January 15, 2015. 2015.
3. Zgibor JC, Orchard TJ, Saul M, et al. Developing and validating a diabetes database in a large health system. *Diab Res Clin Pract* 2007;75:313-9.
4. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care* 2013;36:914-21.
5. Menke A, Orchard TJ, Imperatore G, Mayer-Davis E, Cowie CC. The prevalence of type 1 diabetes in the United States. *Epidemiol* 2013;24:773-4.
6. Pacheco JA, Thomsen LA. Type 2 diabetes mellitus electronic medical record case and control selection algorithms. <https://www.phekborg/sites/phenotype/files/T2DM-algorithmpdf2011>.
7. Alternative methods for health outcomes of interest validation. Available at: [http://mini-sentinel.org/methods/outcome\\_validation/details.aspx?ID=105](http://mini-sentinel.org/methods/outcome_validation/details.aspx?ID=105). 2013. (Accessed August 28, 2015, at [http://mini-sentinel.org/methods/outcome\\_validation/details.aspx?ID=105](http://mini-sentinel.org/methods/outcome_validation/details.aspx?ID=105))
8. O'Connor PJ, Rush WA, Pronk NP, Cherney LM, . Identifying diabetes mellitus or heart disease among health maintenance organization members: Sensitivity, specificity, predictive value, and cost of survey and database methods. *Am J Manag Care* 1998;3:335-42.
9. Saydahi SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. *Am J Epidemiol* 2004;14:507-16.
10. Asghari S, Courteau J, Carpentier AC, Vanasse A. Optimal strategy to identify incidence of diagnostic of diabetes using administrative data. *BMC Medical Research Methodology* 2009;9:62.
11. Harris SB, Glazier RH, Tompkins JW, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. *BMC Health Serv Res* 2010;10:347.
12. Brown JB, Nichols GA, Glauber HS. Case-control study of 10 years of comprehensive diabetes care. *West J Med* 2000;172:85-90.

13. Selby JV, Ray GT, Zhang D, Colby CJ. Excess costs of medical care for patients with diabetes in a managed care population. *Diabetes Care* 1997;20:396-402.
14. Newton KM, Wagner EH, Ramsey SD, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. *J Clin Epidemiol* 1999;52:199-207.
15. Baker AM, Lafata JE, Ward RE, Whithouse F, Divine G. Web-based diabetes care management support system. *Jt Comm J Qual Improv* 2001;27:179-90.
16. Nichols GA, Desai J, Elston LJ, et al. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. *Prev Chronic Dis* 2012;9:E110-Epub2012 Jun7.
17. Raebel MA, Ellis JL, Schroeder EB, et al. Intensification of antihyperglycemic therapy among patients with incident diabetes: a Surveillance Prevention and Management of Diabetes Mellitus (SUPREME-DM) study. *Pharmacoepidemiology and drug safety* 2014;23:699-710.
18. Raebel MA, Xu S, Goodrich GK, et al. Initial antihyperglycemic drug therapy among 241,327 adults with newly identified diabetes from 2005 through 2010: a Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) study. *Ann Pharmacother* 2013;47:1280-91.
19. Nichols GA, Schroeder EB, Karter AJ, et al. Trends in diabetes incidence among 7 million insured adults, 2006-2011: the SUPREME-DM project. *Am J Epidemiol* 2015;181:32-9.
20. Schroeder EB, Powers JD, O'Connor PJ, et al. Prevalence of chronic kidney disease among individuals with diabetes in the SUPREME-DM Project, 2005-2011. *Journal of diabetes and its complications* 2015;29:637-43.
21. Hazelhurst BL, Lawrence JM, Donahoo WT, et al. Automating assessment of lifestyle counseling in electronic health records. *Am J Preventive Med* 2014;46:457-64.
22. Schmittiel JA, Raebel MA, Dyer W, et al. Prescription medication burden in patients with newly diagnosed diabetes: A Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) study. *J Amer Pharm Assoc (2003)* 2014:e251-e9.
23. Xu S, Schroeder EB, Shetterly S, et al. Accuracy of hemoglobin A1c imputation using fasting plasma glucose in diabetes research using electronic health records data. *Stat Optim Inf Comput* 2014;2:93-104.
24. O'Connor PJ, Schmittiel JA, Pathak RD, et al. Randomized trial of telephone outreach to improve medication adherence and metabolic control in adults with diabetes. *Diabetes Care* 2014;37:3317-24.
25. Vasquez-Benitez G, Desai JR, Xu S, et al. Preventable major cardiovascular events associated with uncontrolled glucose, blood pressure, and lipids and active smoking in adults with diabetes with and without cardiovascular disease: a contemporary analysis. *Diabetes Care* 2015;38:905-12.

26. Elston Lafata JE, Karter AJ, O'Connor PJ, et al. Medication adherence does not explain black-white differences in cardiometabolic risk factor control among insured patients with diabetes. *J Gen Intern Med* 2015;Aug 18 [Epub ahead of print]
27. Desai JR, Vazquez-Benitez G, Xu Z, et al. Who must we target now to minimize future cardiovascular events and total mortality? Lessons from the SUPREME-DM cohort study. *Circ Cardiovasc Qual Outcomes* 2015;8:00-.
28. Richesson RL. Data standards in diabetes patient registries. *J Diab Sci Tech* 2011;6:475-85.
29. Ross TR, Ng D, Brown JS, et al. The HMO research network virtual data warehouse: a public data model to support collaboration. *eGEMS*2014:1-8.
30. Engelgau MM, Geiss LS, Manninen DL, et al. Use of services by diabetes patients in managed care organizations. Development of a diabetes surveillance system. CDC Diabetes in Managed Care Work Group. *Diabetes Care*. *Diabetes Care* 1998;21:2062-8.
31. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care* 2004;27 Suppl 2:B10-B21.
32. Mini-Sentinel principles and policies. Available at: [http://www.mini-sentinel.org/about\\_us/principles\\_and\\_policies.aspx](http://www.mini-sentinel.org/about_us/principles_and_policies.aspx). 2014.
33. Nichols G, Desai J, Lawrence J, Reid R, Schroeder EB, Steiner JFea. 5-Year incidence of diabetes among 6.7 million adult HMO members: The SUPREME-DM project. *Diabetes* 2012;61 (Suppl 1):A356.
34. Nichols G, Desai J, Lawrence J, et al. Diabetes prevalence among 7 million adult HMO members, 2005-2009: The SUPREME-DM project. *Diabetes* 2012;61 (Suppl 1):A379.
35. Lo-Ciganic W, Zgibor JC, Ruppert K, Arena VC, Stone RA. Identifying type 1 and type 2 diabetic cases using administrative data: a tree-structured model. *J Diab Sci Tech* 2011;5:486-93.
36. Lawrence JM, Black MH, Zhang JL, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *Am J Epidemiol* 2014;179:27-38.
37. Holt TA, Gunnarsson CL, Cload PA, Rossi F. Identification of undiagnosed diabetes and quality of diabetes care in the United States: cross-sectional study of 11.5 million primary care electronic records. *CMAJ Open* 2013;2:e248-e55.
38. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying identifying cardiovascular and stroke risk factors. *Med Care* 2005;43:480-5.
39. Bobo WV, Cooper WO, Stein CM, et al. Positive predictive value of a case definition for diabetes mellitus using automated administrative health data in children and youth exposed to antipsychotic

- drugs or control medications: a Tennessee Medicaid study. *BMC Medical Research Methodology* 2012;12:128.
40. Rhodes ET, Laffel LMB, Gonzalez TV, Ludwig DS. Accuracy of administrative coding for type 2 diabetes in children, adolescents, and young adults. *Diabetes Care* 2007;30:141-3.
  41. American Diabetes Association. Standards of medical care in diabetes -- 2011. *Diabetes Care* 2011;34:S1-S61.
  42. American Diabetes Association. Standards of medical care in diabetes -- 2015. *Diabetes Care* 2015;38:S1-S94.
  43. JDRF. Type 1 diabetes fact sheet. JDRF2011.
  44. Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014. Atlanta, GA: U S Department of Health and Human Services; 2014.
  45. Zhong VW, Pfaff ER, Beavers DP, et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. *Pediatric Diabetes* 2014;15:573-84.
  46. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2014;37 (Suppl 1):S81-S90.
  47. U.S. Food and Drug Administration. Guidance for Industry: Diabetes Mellitus Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. Available at: <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071627.pdf>. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER); 2008.
  48. Selby JV, Reichman ME, Graham D, et al. A protocol for active surveillance of acute myocardial infarction with use of anti-diabetic agents. Version 5. Available at: [http://www.mini-sentinel.org/work\\_products/Assessments/Mini-Sentinel\\_AMI-and-Anti-Diabetic-Agents\\_Protocol.pdf](http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_AMI-and-Anti-Diabetic-Agents_Protocol.pdf). 2015.
  49. Gerhard T, Raebel MA, Reichman M, et al. Metabolic effects of second generation antipsychotics in youth. Available at: [http://www.mini-sentinel.org/work\\_products/Assessments/Mini-Sentinel\\_Metabolic-Effects-Second-Generation-Antipsychotics-Youth-Report.pdf](http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Metabolic-Effects-Second-Generation-Antipsychotics-Youth-Report.pdf). 2014.
  50. Freeman C, Leonard C, Archdeacon P, Carnahan R, Chrischilles E, Hennessy S. 16 Health Outcomes of Interest for Surveillance Preparedness. Available at: [http://www.mini-sentinel.org/work\\_products/HealthOutcomes/Mini-Sentinel\\_16\\_HOIs\\_Surveillance-Preparedness.pdf](http://www.mini-sentinel.org/work_products/HealthOutcomes/Mini-Sentinel_16_HOIs_Surveillance-Preparedness.pdf). 2014.
  51. Carnahan R, Gagne JJ, Nelson J, et al. Mini-Sentinel prospective routine observational monitoring program tools (PROMPT): rivaroxaban surveillance. Version 2. Available at: [http://www.mini-sentinel.org/work\\_products/Assessments/Mini-Sentinel\\_Rivaroxaban\\_Surveillance.pdf](http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Rivaroxaban_Surveillance.pdf). 2014.

[sentinel.org/work\\_products/assessments/mini-sentinel\\_prompt\\_rivaroxaban-surveillance-plan.Pdf](http://www.mini-sentinel.org/work_products/assessments/mini-sentinel_prompt_rivaroxaban-surveillance-plan.Pdf). 2014.

52. Chrischilles EA, Winiecki SK, Fireman B, et al. Thromboembolic events after immunoglobulin administration. Version 2. Available at: [http://www.mini-sentinel.org/work\\_products/Assessments/Mini-Sentinel\\_Thromboembolic-Events-After-Immunoglobulin-Administration-Protocol.pdf](http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Thromboembolic-Events-After-Immunoglobulin-Administration-Protocol.pdf). 2014.
53. Modular program report. Available at: [http://www.mini-sentinel.org/work\\_products/Assessments/Mini-Sentinel\\_Modular-Program-Report\\_to04mp\\_clinvalidation2\\_Diabetes\\_LabResults.pdf](http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Modular-Program-Report_to04mp_clinvalidation2_Diabetes_LabResults.pdf).
54. Stewart KA, Natzke BM, Williams T, Granger E, Casscells SW, Croghan TW. Temporal trends in antidiabetes drug use in TRICARE following safety warnings in 2007 about rosiglitazone. *Pharmacoepidemiology and drug safety* 2009;18:1048-52.



### XIII. APPENDIX A

#### Published Diabetes Mellitus Algorithms Considered by the Validating Type 1 and Type 2 Diabetes Mellitus in the Mini-Sentinel Distributed Database using the SURveillance, PREvention, and ManagEMENT of Diabetes Mellitus (SUPREME-DM) DataLink Workgroup

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
<b>Any/Either T1DM or T2DM</b>						
Lawrence et al, 2014 <sup>36</sup>	Integrated healthcare system (Kaiser Permanente EHR data)	Any/Either and T1DM vs. T2DM in youth 792,992 youth; 1,568 with DM (<20 years)  Jan-Dec 2009 (12 months)	--Source population was population from which all youth with DM had been previously identified as part of the SEARCH for Diabetes in Youth (SEARCH) Study  -- 1) ICD-9-CM codes for diabetes (250.xx), Rx for insulin and oral antidiabetes meds, and lab test results 2) published DM case identification algorithms with modifications  to these algorithms when data not locally available; and 3) other combinations of indicators  --Investigated total membership and subset with healthcare utilization	T1DM/T2DM ( <i>any DM</i> )  1) individual indicators with highest sensitivity: -- >=1 outpatient (OP) ICD-9-CM Code (250.xx), Sensitivity 94.4% -- Any oral antihyperglycemic or insulin Rx, sensitivity 87.6%  2) Combinations of DM indicators with highest sensitivity: -- >= 1 OP ICD-9-CM code (250.xx) or 1 Rx for insulin, sensitivity 95.9% -- >=1 OP ICD-9-CM code (250.xx) or >=1 inpatient (IP) ICD-9-CM code (250.xx) or 1 Rx for insulin or an oral antihyperglycemic agent, sensitivity 95.0% -- >=1 OP ICD-9-CM code (250.xx) or >=1 IP ICD-9-DM (250.xx), sensitivity 94.8% or multi-indicator algorithm adopted by SUPREME-DM, sensitivity 93.9%  3) PPV highest for >=1 or ICD-9-CM 250.xx, PPV IP, 96.7%; PPV OP, 95.6% --HbA1c >6.5%, PPV 97.6% -- 2 random blood glucose results >=200 mg/dL, PPV 95.6% -- Insulin Rx, PPV 99.8% -- glucagon Rx, PPV 99.4% --Any antidiabetic drug other than metformin, PPV 99.3%	--Assess only youth  --All cases previously validated (SEARCH study)  -- Approach with greatest accuracy in bootstrapping analysis: >=1 OP diagnosis codes or Rx for insulin for both total membership and subset with healthcare utilization, sensitivity 95.9% and PPV 95.5%  -- Short timeframe studied; potential less opportunity for variations in coding than studies of longer duration  --Authors: C-peptide and diabetes autoantibody test results would have assisted in distinguishing T1DM vs T2DM, but these tests not ordered routinely and physicians' assessments of DM type not completely concordant with clinical	Partially.  Available: --ICD-9-CM codes --Rx  Available from some Data Partners: --HbA1c, blood glucose >=200 mg/dl  Strength: Validated  Limitation: Studied only in youth.

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
				<p>T1DM indicators with highest sensitivity:            --OP diagnosis code of T1DM (250x1 or 250.x3), sensitivity 94.8%            --No Rx for metformin, sensitivity 98.0%            -- Rx for insulin or HbA1c &gt;7.5% without metformin Rx, sensitivity 94.6%</p> <p>T2DM indicators with highest sensitivity:            --No OP T1DM diagnosis, sensitivity 93.2%, PPV 81.8%            -- &gt;=1 OP diagnosis of T2DM, sensitivity 86.7%            -- &gt;=1 OP DM diagnosis codes (ICD-9-CM code 250.xx) or an insulin Rx without T1DM diagnosis code, sensitivity 81.8%            -- &gt;=2 T2DM diagnosis codes, sensitivity 71.2%            --Rx for any oral antihyperglycemic, sensitivity 66.5%</p>	assessments of DM type	
Holt et al, 2013 <sup>37</sup>	GE Centricity EHR database; longitudinal records from >9000 primary care clinics (11 million insured patients)	Any/Either (not specified); 11,540,454 (not specifically stated; appears any age was included) "Current to 9/1/2009"	<p>--Feasibility of detecting patients with undiagnosed DM by applying algorithms to EHR data derived from nationally representative sample of US primary care practices</p> <p>--Patients without DM code or DM med use with</p> <p>a) &gt;=2 random glucose &gt;=11.1 mmol/L/ fasting glucose &gt;=7.0</p>	<p>Of 1,110,398 records indicating DM (by code or DM med use), 61.9% contained DM diagnostic code (i.e., 32% identified by medication use only). Of 10 430 056 records for nondiabetic patients, 0.4% (n = 40,359) had &gt;=2 abnormal fasting or random blood glucose values, and 0.2% (n = 23,261) of remaining records had &gt;=1 documented HbA1c &gt;=6.5%</p>	<p>--Considered DM med use as "diagnosed" DM. Excluded those with PCOS.</p> <p>--Total of 1,174,018, of whom 63 620 (5.4%) had "undiagnosed" DM</p> <p>--No internal or external validation</p> <p>--Care could be sought outside of EHR system (e.g., specialty care)</p>	<p>Yes.</p> <p>Available:            --ICD-9-CM codes            --Rx</p> <p>Available from some Data Partners:            --Glucose lab test results</p> <p>Limitations:</p>

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
			mmol/L OR b) HbA1c $\geq$ 6.5% OR c) either of above			Not validated
Zgibor et al, 2007 <sup>3</sup>	Single academic medical center (Pitt)	Any/Either (not specified) 99,144 ( $\geq$ 18 years) Jan 2000 – Dec 2003	--Identified by ICD-9 code 250 from IP, emergency, or OP setting (each considered a separate indicator); any A1c result; blood glucose $>$ 200 mg/dL; or anti-diabetes medication  --Validation studies internally	--Indicators: 1-3) ICD-9 code 250 from IP, emergency, or OP setting; 4) any A1c result (regardless of value); 5) blood glucose $>$ 200 mg/dL; or 6) Rx antidiabetes medication  -- Using two or more indicators or an outpatient diagnosis maximized PPV (96 and 97%) and sensitivity (99 and 100%)	--Internal validation  --Likely a mixture of reasons for HbA1c testing (PPV, sensitivity, and specificity not improved using any specific HbA1c cut point)  --Database does not distinguish T1DM vs. T2DM (authors state this reflects lack of definitive diagnosis in practice and potential recording bias due to coders use of non-specific ICD-9 code)  --Medication data only available for inpatients	Yes.  Available: -- ICD-9-CM codes  Limitation: No external validation  --15-Cohorts WG recommends this indicator-based approach if lab results (HbA1c and/or glucose) available
Solberg et al, 2006 <sup>1</sup>	Not stated; likely one multispecialty network model HMO	Any/Either (not specified) 7148 ("adults") 2000	-- Series of small-sample chart audits on randomly selected cases identified using various criteria  -- Final criteria: $\geq$ 2 outpatient or 1 inpatient DM diagnostic code(s) 250.xx, 357.2, 362.01,	--1) $\geq$ 2 outpatient or 1 inpatient DM diagnostic code(s) 250.xx, 357.2, 362.01, 362.02, 366.41 within a calendar year; or 2) any filled Rx of glucose-lowering medication (not metformin) in the same calendar year  --PPV 96.5% - 100%, sensitivity not reported	--Algorithm to maximize PPV based on prior O'Connor work  Single site; no validation  --Small sample review limits precision of calculated PPV  --Sensitivity of algorithm not determined	Yes.  Available: -- ICD-9-CM codes --Rx  Limitation: No external validation  --15-Cohorts WG recommends this

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
			362.02, 366.41 within a calendar year; or any filled Rx of glucose-lowering medication (not metformin) in the same calendar year			algorithm for identifying cohort with <i>any</i> DM and recommended modification to align with HEDIS measure (modification not evaluated): 1) $\geq 2$ OP or $\geq 1$ emergency department/IP ICD-9 from following, in a given year: 250.XX, 357.2, 362.0X, or 366.41; or 2) Rx for an antidiabetic medication (excluding single-agent metformin) in same year, only if no ICD-9 diagnosis of 251.8 ( <i>other specified disorders of pancreatic internal secretion</i> ), 256.4 ( <i>polycystic ovaries</i> ) or 962.0 ( <i>poisoning by adrenal cortical steroids</i> ) in same/prior year
Birman-Deych et al, 2005 <sup>38</sup>	Medicare Part A claims data	Any/Either 23,657, including 6519 with DM	Cross-sectional study comparing ICD-9-CM data to structured medical record review for beneficiaries with	--ICD-9-DM discharge diagnosis code 250.X in any position, PPV 98%, Sensitivity 75%, NPV 91%	--Population limited to adults with a fib.	Yes. Available: -ICD-9-CM Limitation:

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
		discharge diagnosis code (20-105, mean age 78.8 years (SD, 8.6) April 1998-March 1999	atrial fibrillation; standardized abstraction instruments for 9 cardiovascular and stroke risk factors, including DM; assessed accuracy of ICD-9-CM codes to identify risk factors, chart abstraction gold standard			Population only included adults with a fib
O'Connor et al 1998 <sup>8</sup>	Single am care clinic of staff model HMO	Any/Either (not specified) 3,186 overall, but 1976 ("adult" with both survey and admin data) July 1992 – June 1994	--Combination of survey/self-report, administrative data, and medical record review if survey and database conflicted	--DM identification differed between database and survey methods (sensitivity 91% vs 98%, specificity 99% both methods, PPV 79% vs 81%) -- >= 2 outpatient visits in a 2-year period with any ICD-9 code starting with 250: Sensitivity 91%, specificity 99%, PPV 94% (1883 = no DM; 93 = DM), NPV 99%	--Relatively small Ns; single clinic; no validation --Date range not within that of MSCDM (may be too dated to be useful) --Did not differentiate T1DM vs T2DM --OP codes only --Survey "overreported" DM --14% had no clinic visits within one year period (i.e., no information available to classify DM status)	Yes. Available: --ICD-9-CM codes Limitations: No validation; dated; OP codes only
<b>T1DM vs. T2DM, T1DM Only, or T2DM Only</b>						
Klompas et al, 2013	Multi-specialty am care	T1DM vs. T2DM	-- >= two ICD-9 codes -- T1DM: Ratio of	T1DM using ICD-9-CM Codes only: a) >=2 ICD-9 codes for T1DM only (250.X1 or	--ICD-9 codes, lab and Rx data	Partially.

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
4	practice (Atrius Health EHR data)	43,177 (0 - >80 years) June 2006 – Sept 2010	T1/T2 codes >0.5 and glucagon Rx; ratio of T1/T2 codes >0.5 and no oral hypoglycemic other than metformin; C-peptide negative; autoantibodies present; urine acetone test strips Rx; otherwise T2  --Multiple candidate algorithms presented  --Chart review confirmed 298 of 310  --Validated in same system  --About 13,000 not classified  --Sensitivity and PPV of various algorithms for T1DM	250.X3): Sensitivity 26%, PPV 94%  b) >=2 ICD-9 codes for T1DM and any number of T2DM (250.X0 or 250.X2) codes: Sensitivity 90%, PPV 57%  --T1DM requiring ≥2 ICD-9 diagnoses for 250.X1 or 250.X3 + a current Rx for insulin + no Rx for an oral antidiabetic agent at any time (excluding metformin): Sensitivity 32%, PPV 81%  --Algorithm that maximized PPV while maintaining acceptable sensitivity: Ratio of T1DM to T2DM >0.5 + no prescription for an oral antidiabetic drug (excluding metformin): PPV = 96%, sensitivity = 61%  T1DM of optimized algorithm; any of following: <sup>b</sup> a) Ratio of T1DM to T2DM codes >0.5 and glucagon Rx or b) Ratio of T1DM to T2DM codes >0.5 and no record of an oral hypoglycemic other than metformin or c) negative C-peptide or d) pos. DM autoantibodies or e) urine acetone test strips Rx  Raw Sensitivity 97%, population-weighted Sensitivity 65%, population-weighted PPV 88%	--Focused optimized algorithm on sensitivity to T1DM because even slight misclassification of T1DM as T2DM is substantially magnified after weighting for greater size of T2DM population  --Multiple algorithms tested; maximizing PPV often at cost of sensitivity and vice versa  --Persistent sources of error in the algorithms: physician miscoding, free text recording, EHR miscoding, algorithm programming errors	Available: --ICD-9-CM codes --Rx  Available from some Data Partners: --Glucose lab test results  Not available: --C-peptide --DM auto-antibodies  Strength: Multiple algorithms tested.  All ages included.  --Algorithm with maximized PPV and acceptable sensitivity recommended by "15-Cohorts" WG for T1DM (optimized algorithm recommended only if lab test data available in MSDD)

Figure 5. Algorithm for identifying T2DM cases in the EMR

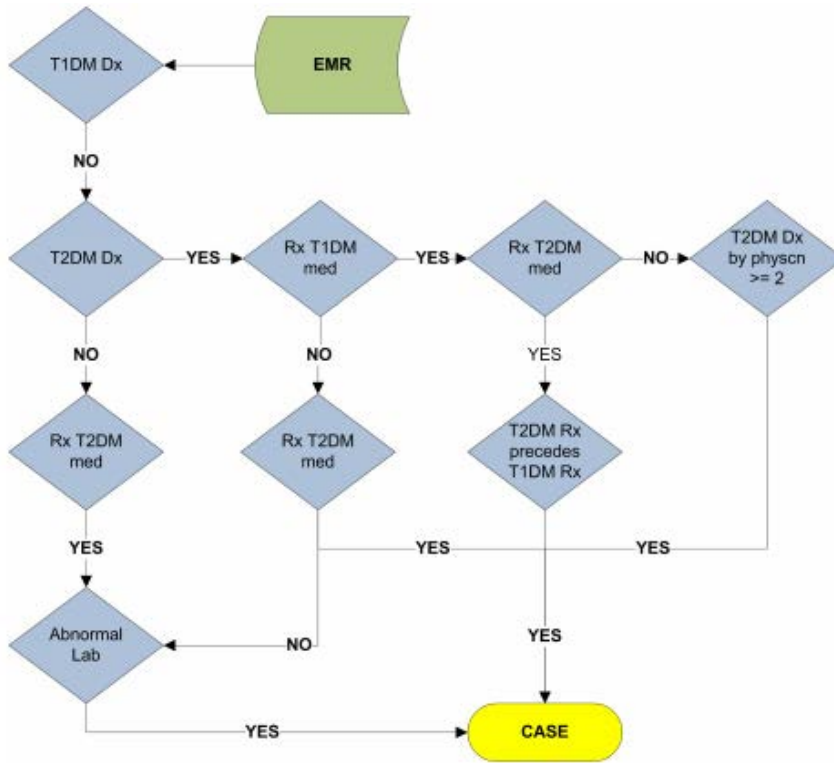
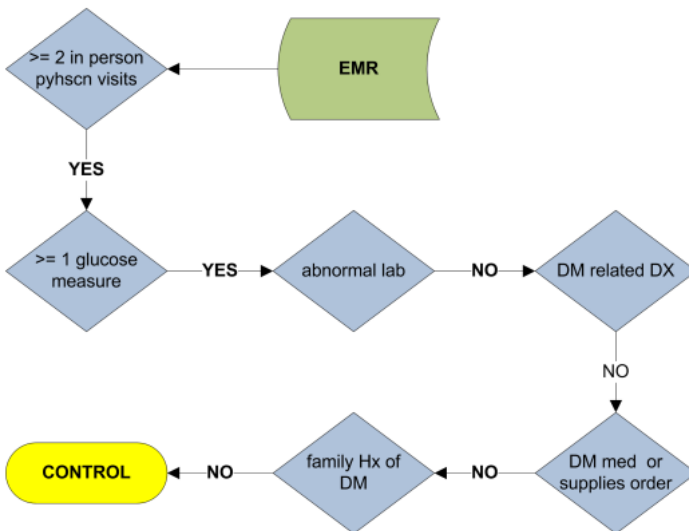


Figure 6. Algorithm for identifying T2DM controls in the EMR





Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
Pacheco and Thompson (eMERGE D2DM algorithm) 2011 <sup>6</sup>	Only algorithm presented, no actual validation	T2DM Age and timeframe not given	--T2DM CASE selection using ICD-9 codes, T1DM meds (insulin, symlin) order or dates, T2DM meds order or dates, labs (FBS, RBS, A1c) at least maximum value  --T2DM CONTROL selection using ICD-9 codes, T1DM meds (insulin, symlin) order or dates, T2DM meds order or dates, labs (FBS, RBS, A1c) at least maximum value, Family Hx DM, DM medical supply orders, count of dates patient had face to face outpatient clinic encounter	<b>Error! Not a valid result for table.</b>  <b>Figure 6</b>	--does not call out metformin which is not specific for T2 DM	Partially Available: ICD-9 codes Rx OP visits Available from some Data Partners: Glucose lab test results  Not available: Family history DM  Limitation: Did not remove metformin only from case selection criteria
Bobo et al, 2012 <sup>39</sup>	Tennessee Medicaid; youth starting an atypical antipsychotic or control medication	T1DM and T2DM 172,014 of whom 64 met DM case definition (6-24 years) Jan 1996 – Dec 2007	--T1DM or T2DM 1) Primary IP diagnosis (250, 250.0X, 250.1X, 250.2X, 250.3X, or 250.9X; 2) >=2 other encounters of different types (e.g., an OP diagnosis with Rx); or 3) IP stay with secondary discharge diagnosis for one of above codes + confirmatory anti-DM Rx or additional DM code within 120 days --PCOS (265.4) excluded --T1DM defined by insulin	--T1DM PPV: 80.0% T2DM and unspecified DM combined PPV: 83.9% (PPV 74.2% if only those coded as T2DM were included) --Estimated sensitivity of the definition, based on adjudication for a sample of 30 cases not meeting the automated database definition: 64.8%	--Developed and validated in same population; validated in convenience sample --Single oral agent Rx did not meet an exclusion criterion because such drugs may be occasionally prescribed while awaiting confirmatory testing results for T1DM --Population was small number of children, adolescents, young adults --Most common source of	Yes. Available: -- ICD-9-CM codes --Rx  Limitation: Studied in population 6 – 24 years.

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
			Rx with <=1 oral hypoglycemic Rx; other cases considered T2DM --Medical records reviewed/adjudicated for cases that met automated database definition as well as for a sample of persons with other DM care encounters		misclassification was subthreshold hyperglycemia (6% and 14% of adjudicated cases that respectively met primary or secondary case definitions)	
Lo-Ciganic, 2011 <sup>35</sup>	Single academic center (Pitt)	T1DM vs. T2DM 129,684 (>=18 years) Jan 2000 – Sept 2009	--Clinical criteria applied to administrative data Nonparametric classification tree models (classification and regression trees [CART]) fit/built a tree structured model to distinguish between T1DM and T2DM using admin data from Pitt Medical Center database/EHR --Randomly selected medical records reviewed to verify	Main predictors distinguishing T1 from T2: --ICD-9 codes: a) for T1DM [250.x1 or 250.x3] without codes for T2DM [250.x0 or 250.x2], b) for T1DM and T2DM --Age < 40 --Inpatient insulin use --Inpatient oral hypoglycemic use --Episode of diabetic ketoacidosis --For T1DM: Sensitivity 92.8%, specificity 99.3%, PPV 89.5%, NPV 99.5%	--No internal or external validation --Requires inpatient medication data --79,963 patients excluded for “no strong indicators” of either DM type or missing data	Partially. Available: -- ICD-9-CM codes Not available: --Inpatient medication use  Limitations: No validation. Adults only.
Rhodes et al, 2007 <sup>40</sup>	Specialty clinics, Boston Children’s Hospital	T2DM (based on single ICD-9-CM code) 432 (<26 years) July 2003 – January 2005 (30 months)	--Chart review -- >= one ICD-9-CM T2DM code (250.X0 or 250.X2, X=0–9); inpatient/outpatient sites -- >= one ICD-9-CM T1DM code (250,X1 or 250.X3) -- PPV only (sensitivity	T1DM: >= one ICD-9-CM T1DM code (250,X1 or 250.X3); PPV 97.0%  T2DM: >= one ICD-9-CM T2DM code (250.X0 or 250.X2; PPV 16.0%	-- Single system; no validation --Sensitivity not reported --T2DM codes include “unspecified” DM  --Given a) increasing pediatric obesity, b) patients with phenotypic T2DM may have pancreatic autoimmunity, and c) African	Yes.  Available: -- ICD-9-CM codes  Limitations: Not validated. Age < 26 only  --15-Cohorts WG recommended

Lead Author, year	Setting	T1DM or T2DM; N (age range); Timeframe	Methods	Results	Comments	Applicable to MSCDM?
			not reported)		Americans may present with nonautoimmune (idiopathic) T1(b)DM, differentiating T1/T2DM at diagnosis may be difficult/coded as unspecified	T1DM algorithm if FDA interested in identifying cohort of pediatric, adolescent, and young adults with T1DM

## XIV. APPENDIX B

**Table 1. ICD-9-CM Code Definitions**

Codes for Inclusion	
250.xx	Diabetes Mellitus
250.x1, 250.x3	Type 1 Diabetes Mellitus
250.x0, 250.x2	Type 2 Diabetes Mellitus
357.2	Diabetic neuropathy
362.01	Background Diabetic Retinopathy
362.02	Proliferative Diabetic Retinopathy
362.03	Nonproliferative Diabetic Retinopathy
362.04	Mild Nonproliferative Diabetic Retinopathy
362.05	Moderate Nonproliferative Diabetic Retinopathy
362.06	Severe Nonproliferative Diabetic Retinopathy
362.07	Diabetic Macular Edema
366.41	Diabetic cataract
Codes for Exclusion	
251.8	Other specified disorders of pancreatic internal secretion
256.4	Polycystic ovaries
962.0	Poisoning by adrenal cortical steroids
250.10 (exclusion for Type 2 only)	Diabetes with ketoacidosis, Type II or unspecified type, not stated as uncontrolled
250.12 (exclusion for Type 2 only)	Diabetes with ketoacidosis, Type II or unspecified type, uncontrolled

As can be seen in Table 2, the algorithms we tested either did not provide a specific list of diabetes medications they included (e.g., Solberg) or the medication lists used were out of date (e.g., Zgibor). Therefore, we applied the updated SUPREME-DM medication list when testing all algorithms. The SUPREME-DM medication list was updated through June 2014, the end date of the data linkage we examined. For example, canagliflozin is included in the SUPREME-DM updated medication list, but is not included in any of the other algorithms because it is too newly marketed.

**Table 2. Antidiabetic Medication Lists from the Original Algorithms**

Solberg (2006) <sup>1</sup>	Zgibor (2007) <sup>3</sup>	SUPREME-DM <sup>16</sup>	Klompas (2013) <sup>4</sup> – “oral agents”	eMERGE Type 2DM medications (2011) <sup>6</sup>
Not defined	Acarbose	Acarbose		Acarbose
	Acetohexamide	Acetohexamide		Acetohexamide
		Alogliptin <sup>e</sup>		
		Canagliflozin		
	Chlorpropramide	Chlorpropramide		Chlorpropramide
		Dapagliflozin <sup>e</sup>		
		Empagliflozin <sup>e</sup>		
		Exenatide	Exenatide <sup>a</sup>	Exenatide

Solberg (2006) <sup>1</sup>	Zgibor (2007) <sup>3</sup>	SUPREME-DM <sup>16</sup>	Klompas (2013) <sup>4</sup> – “oral agents”	eMERGE Type 2DM medications (2011) <sup>6</sup>
			Gliclazide <sup>b</sup>	
	Glimepiride	Glimepiride	Glimepiride	Glimepiride
	Glipizide	Glipizide	Glipizide	Glipizide
	Glucagon <sup>c</sup>			
	Glyburide	Glyburide	Glyburide	Glyburide
	Insulin	Insulin		
		Linagliptin		
		Liraglutide		
	Metformin	Metformin		Metformin
	Miglitol	Miglitol		Miglitol
		Nateglinide	Nateglinide	Nateglinide
	Pioglitazone	Pioglitazone	Pioglitazone	Pioglitazone
		Pramlintide	Pramlintide <sup>a</sup>	
	Repaglinide	Repaglinide	Repaglinide	Repaglinide
	Rosiglitazone	Rosiglitazone	Rosiglitazone	Rosiglitazone
		Saxagliptin		
		Sitagliptin	Sitagliptin	Sitagliptin
	Tolazamide	Tolazamide		Tolazamide
	Tolbutamide	Tolbutamide		
	Troglitazone <sup>d</sup>			Troglitazone <sup>d</sup>

<sup>a</sup> Not oral but included in Klompas algorithm listing of “oral agents.”

<sup>b</sup> Not marketed in US; will not be included in any/either DM or for T1DM or T2DM.

<sup>c</sup> Will not be included in tested algorithms for any/either DM or for T1DM or T2DM.

<sup>d</sup> Removed from US market on 3/21/2000. Will not be included in tested algorithms for any/either DM or for T1DM or T2DM.

<sup>e</sup> Included in this listing for future application. Not included in SUPREME-DM DataLink refresh through June 2014.